Nonparametric kernel estimation of the probability density function of regression errors using estimated residuals

By Rawane SAMB

Université Pierre et Marie Curie, LSTA.

Abstract

This paper deals with the nonparametric density estimation of the regression error term assuming its independence with the covariate. The difference between the feasible estimator which uses the estimated residuals and the unfeasible one using the true residuals is studied. An optimal choice of the bandwidth used to estimate the residuals is given. We also study the asymptotic normality of the feasible kernel estimator and its rate-optimality.

Keywords: Kernel density estimation, Leave-one-out kernel estimator, Two-steps estimator.

1 Introduction

Consider a sample $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ of independent and identically distributed (i.i.d) random variables, where Y is the univariate dependent variable and the covariate X is of dimension d. Let $m(\cdot)$ be the conditional expectation of Y given X and let ε be the related regression error term, so that the regression error model is

$$Y_i = m(X_i) + \varepsilon_i, \quad i = 1, \ldots, n. \tag{1.1}$$

We wish to estimate the probability distribution function (p.d.f) of the regression error term, $f(\cdot)$, using the nonparametric residuals. Our potential applications are as follows. First, an estimation of the p.d.f of ε is an important tool for understanding the residuals behavior and therefore the fit of the regression model (1.1). This estimation of $f(\cdot)$ can be used for goodness-of-fit tests of a specified error distribution in a parametric regression setting. Some examples can be founded in Loynes (1980), Akritas and Van Keilegom (2001), Cheng and Sun (2008). The estimation of the density of the regression error term can also be useful for testing the symmetry of the residuals. See Ahmad et Li (1997), Dette et *al.* (2002). Another interest of the estimation of f is that it can be used for constructing nonparametric estimators for the density and hazard function of Y given X, as related in Van Keilegom and Veraverbeke (2002). This estimation of f is also important when

are interested in the estimation of the p.d.f of the response variable Y. See Escanciano and Jacho-Chavez (2010). Note also that an estimation of the p.d.f of the regression errors can be useful for proposing a mode forecast of Y given $X = x$. This mode forecast is based on an estimation of $m(x) + \arg\min_{\epsilon \in \mathbb{R}} f(\epsilon)$.

Relatively little is known about the nonparametric estimation of the p.d.f and the cumulative distribution function (c.d.f) of the regression error. Up to few exceptions, the nonparametric literature focuses on studying the distribution of Y given X. See Roussas (1967, 1991), Youndjé (1996) and references therein. Akritas and Van Keilegom (2001) estimate the cumulative distribution function of the regression error in heteroscedastic model. The estimator proposed by these authors is based on a nonparametric estimation of the residuals. Their result show the impact of the estimation of the residuals on the limit distribution of the underlying estimator of the cumulative distribution function. Müller, Schick and Wefelmeyer (2004) consider the estimation of moments of the regression error. Quite surprisingly, under appropriate conditions, the estimator based on the true errors is less efficient than the estimator which uses the nonparametric estimated residuals. The reason is that the latter estimator better uses the fact that the regression error ε has mean zero. Efromovich (2005) consider adaptive estimation of the p.d.f of the regression error. He gives a nonparametric estimator based on the estimated residuals, for which the Mean Integrated Squared Error (MISE) attains the minimax rate. Fu and Yang (2008) study the asymptotic normality of the estimators of the regression error p.d.f in nonlinear autoregressive models. Cheng (2005) establishes the asymptotic normality of an estimator of $f(\cdot)$ based on the estimated residuals. This estimator is constructed by splitting the sample into two parts: the first part is used for the construction of estimator of $f(\cdot)$, while the second part of the sample is used for the estimation of the residuals.

The focus of this paper is to estimate the p.d.f of the regression error using the estimated residuals, under the assumption that the covariate X and the regression error ε are independent. In a such setup, it would be unwise to use a conditional approach based on the fact that $f(\epsilon) = f(\epsilon|x) = \varphi\left(m(x) + \epsilon|x\right)$, where $\varphi(\cdot|x)$ is the p.d.f of Y given $X = x$. Indeed, the estimation of $m(\cdot)$ and $\varphi(\cdot|x)$ are affected by the curse of dimensionality, so that the resulting estimator of $f(\cdot)$ would have considerably a slow rate of convergence if the dimension of X is high. The approach proposed here uses a two-steps procedure which, in a first step, replaces the unobserved regression error terms by some nonparametric estimator $\widehat{\varepsilon}_i$. In a second step, the estimated $\widehat{\varepsilon}_i$'s are used to estimate nonparametrically $f(\cdot)$, as if they were the true ε_i's. If proceeding so can circumvent the curse of dimensionality, a challenging issue is to evaluate the impact of the estimated residuals on the final estimator of $f(\cdot)$. Hence one of the contributions of our study is to analyze the effect

of the estimation of the residuals on the regression errors p.d.f. Kernel estimators. Next, an optimal choice of the bandwidth used to estimate the residuals is given. Finally, we study the asymptotic normality of the feasible Kernel estimator and its rate-optimality.

The rest of this paper is organized as follows. Section 2 presents ours estimators and proposes an asymptotic normality of the (naive) conditional estimator of the density of the regression error term. Sections 3 and 4 group our assumptions and main results. The conclusion of this chapter is given in Section 5, while the proofs of our results are gathered in section 6 and in an appendix.

2 Some nonparametric conditional estimator of the density of the regression error

To illustrate the potential impact of the dimension d of the X_i's, let us first consider a naive conditional estimator of the p.d.f $f(\cdot)$ of the regression error term ε. Let $\varphi(\cdot|x)$ and $f(\cdot|x)$ be respectively the p.d.f. of Y and ε given $X = x$. Since $f(\epsilon|x) = \varphi(m(x) + \epsilon|x)$, using the independence of X and ε gives

$$f(\epsilon) = f(\epsilon|x) = \varphi\left(m(x) + \epsilon|x\right). \tag{2.1}$$

Consider some Kernel functions K_0, K_1 and some bandwidths b_0, h_0 and h_1. The expression (2.1) of f suggests to use the Kernel nonparametric estimator

$$\widetilde{f}_n(\epsilon|x) = \frac{\frac{1}{nh_0^d h_1}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right) K_1\left(\frac{Y_i - \widehat{m}_n(x) - \epsilon}{h_1}\right)}{\frac{1}{nh_0^d}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right)},$$

where $\widehat{m}_n(x)$ is the Nadaraya-Watson (1964) estimator of $m(x)$ defined as

$$\widehat{m}_n(x) = \frac{\sum_{j=1}^n Y_j K_0\left(\frac{X_j - x}{b_0}\right)}{\sum_{j=1}^n K_0\left(\frac{X_j - x}{b_0}\right)}. \tag{2.2}$$

The first result presented in this chapter is the following proposition.

proposition 2.1. *Define*

$$\mu_1(x,\epsilon) = \frac{\partial^2 \varphi\left(x, m(x) + \epsilon\right)}{\partial^2 x}\int z K_0(z) z^\top dz, \quad \mu_2(x,\epsilon) = \frac{\partial^2 \varphi\left(x, m(x) + \epsilon\right)}{\partial^2 y}\int v^2 K_1(v) dv,$$

and suppose that h_0 *decrease to* 0 *such that* $nh_0^{2d}/\ln n \to \infty$, $\ln(1/h_0)/\ln(\ln n) \to \infty$ *and*

$$(\mathbf{A_0}): \quad nh_0^d h_1 \to \infty, \quad \left(\frac{nh_0^d}{h_1}\right)\left(b_0^4 + \frac{\ln n}{nb_0^d}\right) = o(1),$$

when $n \to \infty$. Then under Assumptions $(A_1) - (A_{10})$ given in the next section, we have

$$\sqrt{nh_0^d h_1}\left(\widetilde{f}_n(\epsilon|x) - \overline{\widetilde{f}}_n(\epsilon|x)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(\epsilon|x)}{g(x)} \int\int K_0^2(z)K_1^2(v)dzdv\right),$$

where $g(\cdot)$ is the marginal density of X and

$$\overline{\widetilde{f}}_n(\epsilon|x) = f(\epsilon|x) + \frac{h_0^2\mu_1(x,\epsilon)}{2g(x)} + \frac{h_1^2\mu_2(x,\epsilon)}{2g(x)} + o\left(h_0^2 + h_1^2\right).$$

This results suggests that an optimal choice of the bandwidths h_0 and h_1 should achieve the minimum of the asymptotic mean square expansion first order terms

$$AMSE\left(\widetilde{f}_n(\epsilon|x)\right) = \left[\frac{h_0^2\mu_1(x,\epsilon)}{2g(x)} + \frac{h_1^2\mu_2(x,\epsilon)}{2g(x)}\right]^2 + \frac{f(\epsilon|x)\int K_0^2(z)dz\int K_1^2(v)dv}{nh_0^d h_1 g(x)}.$$

Elementary calculations yield that the resulting optimal bandwidths h_0 and h_1 are all proportional to $n^{-1/(d+5)}$, leading to the exact consistency rate $n^{-2/(d+5)}$ for $\widetilde{f}_n(x|\epsilon)$. In the case $d = 1$, this rate is $n^{-1/3}$, which is worst than the rate $n^{-2/5}$ achieved by the optimal Kernel estimator of an univariate density. See Bosq and Lecoutre (1987), Scott (1992), Wand and Jones (1995). Note also that the exponent $2/(d+5)$ decreases to 0 with the dimension d. This indicates a negative impact of the dimension d on the performance of the estimator, the so-called curse of dimensionality. The fact that $\widetilde{f}_n(\epsilon|x)$ is affected by the curse of dimensionality is a consequence of conditioning. Indeed, (2.1) identifies the unconditional $f(\epsilon)$ with the conditional distribution of the regression error given the covariate.

To avoid this curse of dimensionality in the nonparametric kernel estimation of $f(\epsilon)$, our approach proposed here builds, in a first step, the estimated residuals

$$\widehat{\varepsilon}_i = Y_i - \widehat{m}_{in}, \quad i = 1, \ldots, n, \tag{2.3}$$

where $\widehat{m}_{in} = \widehat{m}_{in}(X_i)$ is a leave-one out version of the Kernel regression estimator (2.2),

$$\widehat{m}_{in} = \frac{\sum_{\substack{j=1\\j\neq i}}^n Y_j K_0\left(\frac{X_j - X_i}{b_0}\right)}{\sum_{\substack{j=1\\j\neq i}}^n K_0\left(\frac{X_j - X_i}{b_0}\right)}. \tag{2.4}$$

It is tempting to use, in a second step, the estimated $\widehat{\varepsilon}_i$ as if they were the true residuals ε_i. This would ignore that the $\widehat{m}_n(X_i)$'s can deliver severely biased estimations of the $m(X_i)$'s for those X_i which are close to the boundaries of the support $\mathcal{X}$ of the covariate distribution. To that aim, our proposed estimator trims

the observations X_i outside an inner subset $\mathcal{X}_0$ of $\mathcal{X}$,

$$\widehat{f}_{1n}(\epsilon) = \frac{1}{b_1 \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right)} \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right) K_1\left(\frac{\widehat{\varepsilon}_i - \epsilon}{b_1}\right). \tag{2.5}$$

This estimator is the so-called two-steps Kernel estimator of $f(\epsilon)$. In principle, it would be possible to assume that $\mathcal{X}_0$ grows to $\mathcal{X}$ with a negligible rate compared to the bandwidth b_1. This would give an estimator close to the more natural Kernel estimator $\sum_{i=1}^n K\left((\widehat{\varepsilon}_i - \epsilon)/b_1\right)/(nb_1)$. However, in the rest of the paper, a fixed subset $\mathcal{X}_0$ will be considered for the sake of simplicity.

Observe that the two steps Kernel estimator $\widehat{f}_{1n}(\epsilon)$ is a feasible estimator in the sense that it does not depend on any unknown quantity, as desirable in practice. This contrasts with the unfeasible ideal Kernel estimator

$$\widetilde{f}_{1n}(\epsilon) = \frac{1}{b_1 \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right)} \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right) K_1\left(\frac{\varepsilon_i - \epsilon}{b_1}\right), \tag{2.6}$$

which depends in particular on the unknown regression error terms. It is however intuitively clear that $\widehat{f}_{1n}(\epsilon)$ and $\widetilde{f}_{1n}(\epsilon)$ should be closed, as illustrated by the results of the next section.

3 Assumptions

The following assumptions are used in our mains results.

$(\mathbf{A_1})$ *The support $\mathcal{X}$ of X is a compact subset of $\mathbb{R}^d$ and $\mathcal{X}_0$ is an inner closed subset of $\mathcal{X}$ with non empty interior,*

$(\mathbf{A_2})$ *the p.d.f. $g(\cdot)$ of the i.i.d. covariates X, X_i is strictly positive over $\mathcal{X}_0$, and has continuous second order partial derivatives over $\mathcal{X}$,*

$(\mathbf{A_3})$ *the regression function $m(\cdot)$ has continuous second order partial derivatives over $\mathcal{X}$,*

$(\mathbf{A_4})$ *the i.i.d. centered error regression terms $\varepsilon, \varepsilon_i$'s, have finite 6th moments, and are independent of the covariates X, X_i's,*

$(\mathbf{A_5})$ *the probability density function $f(\cdot)$ has bounded continuous second order derivatives over $\mathbb{R}$ and satisfies, for $h_p(e) = e^p f(e)$, $\sup_{e\in\mathbb{R}} |h_p^{(k)}(e)| < \infty$, $p \in [0,2]$, $k \in [0,2]$,*

$(\mathbf{A_6})$ *the p.d.f φ of (X, Y) has bounded continuous second order partial derivatives over $\mathbb{R}^d \times \mathbb{R}$,*

$(\mathbf{A_7})$ *the Kernel function K_0 is symmetric, continuous over $\mathbb{R}^d$ with support contained in $[-1/2, 1/2]^d$ and $\int K_0(z)dz = 1$,*

($\mathbf{A_8}$) *the Kernel function* K_1 *has a compact support, is three times continuously differentiable over* $\mathbb{R}$*, and satisfies* $\int K_1(v)dv = 1$ *and* $\int vK_1(v)dv = 0$*,*

($\mathbf{A_9}$) *the bandwidth* b_0 *decreases to* 0 *and satisfies, for* $d^* = \sup\{d+2, 2d\}$, $nb_0^{d^*}/\ln n \to \infty$ *and* $\ln(1/b_0)/\ln(\ln n) \to \infty$ *when* $n \to \infty$*,*

($\mathbf{A_{10}}$) *the bandwidth* b_1 *decreases to* 0 *and satisfies* $n^{(d+8)}b_1^{7(d+4)} \to \infty$ *when* $n \to \infty$*.*

Assumptions (A_2), (A_3), (A_5) and (A_6) impose that all the functions to be estimated nonparametrically have two bounded derivatives. Consequently the conditions $\int zK_0(z)dz = 0$ and $\int vK_1(v)dv = 0$, as assumed in (A_7) and (A_8), represent standard conditions ensuring that the bias of the resulting nonparametric estimators (2.2) and (2.6) are of order b_0^2 and b_1^2. Assumption (A_4) states independence between the regression error terms and the covariates, which is the main condition for (2.1) to hold. The differentiability of K_1 imposed in (A_8) is more specific to our two-steps estimation method. Assumption (A_8) is used to expand the two-steps Kernel estimator $\widehat{f}_{1n}$ in (2.5) around the unfeasible one $\widetilde{f}_{1n}$ from (2.6), using the residual error estimation $\widehat{\varepsilon}_i - \varepsilon_i$'s and the derivatives of K_1 up to third order. Assumption (A_9) is useful for obtaining the uniform convergence of the regression estimator $\widehat{m}_n$ defined in (2.2) (see for instance Einmahl and Mason, 2005), and also gives a similar consistency result for the leave-one-out estimator $\widehat{m}_{in}$ in (2.4). Assumption (A_{10}) is needed in the study of the difference between the feasible estimator $\widehat{f}_{1n}$ and the unfeasible estimator $\widetilde{f}_{1n}$.

4 Main results

This section is devoted to our main results. The first result we give here concerns the pointwise consistency of the nonparamatric Kernel estimator $\widehat{f}_{1n}$ of the density f. Next, the optimal first-step and second-step bandwidths used to estimated f are proposed. We finish this section by establishing an asymptotic normality for the estimator $\widehat{f}_{1n}$.

4.1 Pointwise weak consistency

The next result gives the order of the difference between the feasible estimator and the theoretical density of the regression error at a fixed point ϵ.

Theorem 4.1. *Under* $(A_1) - (A_5)$ *and* $(A_7) - (A_{10})$*, we have, when* b_0 *and* b_1 *go to* 0*,*

$$\widehat{f}_{1n}(\epsilon) - f(\epsilon) = O_{\mathbb{P}}\Big(AMSE(b_1) + R_n(b_0, b_1)\Big)^{1/2},$$

where

$$AMSE(b_1) = \mathbb{E}_n\left[\left(\widetilde{f}_{1n}(\epsilon) - f(\epsilon)\right)^2\right] = O_{\mathbb{P}}\left(b_1^4 + \frac{1}{nb_1}\right),$$

and

$$R_n(b_0, b_1) = b_0^4 + \left[\frac{1}{(nb_1^5)^{1/2}} + \left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]^2\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 + \left[\frac{1}{b_1} + \left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]^2\left(b_0^4 + \frac{1}{nb_0^d}\right)^3.$$

The result of Theorem 4.1 is based on the evaluation of the difference between $\widehat{f}_{1n}(\epsilon)$ and $\widetilde{f}_{1n}(\epsilon)$. This evaluation gives an indication about the impact of the estimation of the residuals on the nonparametric estimation of the regression error density.

4.2 Optimal first-step and second-step bandwidths for the pointwise weak consistency

As shown in the next result, Theorem 4.2 gives some guidelines for the choice of the optimal bandwidth b_0 used in the nonparametric regression errors estimation. As far as we know, the choice of an optimal b_0 has not been addressed before. In what follows, $a_n \asymp b_n$ means that $a_n = O(b_n)$ and $b_n = O(a_n)$, i.e. that there is a constant $C > 0$ such that $|a_n|/C \leq |b_n| \leq C|a_n|$ for n large enough.

Theorem 4.2. *Suppose that $(A_1) - (A_5)$ and $(A_7) - (A_{10})$ are satisfied, and define*

$$b_0^* = b_0^*(b_1) = \arg\min_{b_0} R_n(b_0, b_1).$$

where the minimization is performed over bandwidth b_0 fulfilling (A_9). Then the bandwidth b_0^* satisfies

$$b_0^* \asymp \max\left\{\left(\frac{1}{n^2b_1^3}\right)^{\frac{1}{d+4}}, \left(\frac{1}{n^3b_1^7}\right)^{\frac{1}{2d+4}}\right\},$$

and we have

$$R_n(b_0^*, b_1) \asymp \max\left\{\left(\frac{1}{n^2b_1^3}\right)^{\frac{4}{d+4}}, \left(\frac{1}{n^3b_1^7}\right)^{\frac{4}{2d+4}}\right\}.$$

Our next theorem gives the conditions for which the estimator $\widehat{f}_{1n}(\epsilon)$ reaches the optimal rate $n^{-2/5}$ when b_0 takes the value b_0^*. We prove that for $d \leq 2$, the bandwidth that minimizes the term $AMSE(b_1) + R_n(b_0^*, b_1)$ has the same order as $n^{-1/5}$, yielding the optimal order $n^{-2/5}$ for $(AMSE(b_1) + R_n(b_0^*, b_1))^{1/2}$.

Theorem 4.3. *Assume that* $(A_1) - (A_5)$ *and* $(A_7) - (A_{10})$ *are satisfied, and set*

$$b_1^* = \arg\min_{b_1}\Big(AMSE(b_1) + R_n(b_0^*, b_1)\Big),$$

where $b_0^* = b_0^*(b_1)$ *is defined as in Theorem 4.2. Then*

i. For $d \leq 2$, *the bandwidth* b_1^* *satisfies*

$$b_1^* \asymp \left(\frac{1}{n}\right)^{\frac{1}{5}},$$

and we have

$$\Big(AMSE(b_1^*) + R_n(b_0^*, b_1^*)\Big)^{\frac{1}{2}} \asymp \left(\frac{1}{n}\right)^{\frac{2}{5}}.$$

ii. For $d \geq 3$, b_1^* *satisfies*

$$b_1^* \asymp \left(\frac{1}{n}\right)^{\frac{3}{2d+11}},$$

and we have

$$\Big(AMSE(b_1^*) + R_n(b_0^*, b_1^*)\Big)^{\frac{1}{2}} \asymp \left(\frac{1}{n}\right)^{\frac{6}{2d+11}}.$$

The results of Theorem 4.3 show that the rate $n^{-2/5}$ is reachable if and only when $d \leq 2$. These results are derived from Theorem 4.2. This latter indicates that if b_1 is proportional to $n^{-1/5}$, the bandwidth b_0^* has the same order as

$$\max\left\{\left(\frac{1}{n}\right)^{\frac{7}{5(d+4)}}, \left(\frac{1}{n}\right)^{\frac{8}{5(2d+4)}}\right\} = \left(\frac{1}{n}\right)^{\frac{8}{5(2d+4)}}.$$

For $d \leq 2$, this order of b_0^* is smaller than the one of the optimal bandwidth b_{0*} obtained for pointwise or mean square estimation of $m(\cdot)$ using a Kernel estimator. In fact, it has been shown in Nadaraya (1989, Chapter 4) that the optimal bandwidth b_{0*} for estimating $m(\cdot)$ is obtained by minimizing the order of the risk function

$$r_n(b_0) = \mathbb{E}\left[\int \mathbb{1}\,(x \in \mathcal{X})\,(\widehat{m}_n(x) - m(x))^2\,\widehat{g}_n^2(x) w(x) dx\right],$$

where $\widehat{g}_n(x)$ is a nonparametric Kernel estimator of $g(x)$, and $w(\cdot)$ is a nonnegative weight function, which is bounded and squared integrable on $\mathcal{X}$. If $g(\cdot)$ and $m(\cdot)$ have continuous second order partial derivatives over their supports, Nadaraya (1989, Chapter 4) shows that $r_n(b_0)$ has the same order as $b_0^4 + \big(1/(nb_0^d)\big)$, leading to the optimal bandwidth $\widehat{b}_0 = n^{-1/(d+4)}$ for the convergence of the estimator $\widehat{m}_n(\cdot)$ of $m(\cdot)$ in the set of the square integrable functions on $\mathcal{X}$.

For d=1, the optimal order of b_0^* is $n^{-(1/5)\times(4/3)}$ which goes to 0 slightly faster than $n^{-1/5}$, the optimal order of the bandwidth $\widehat{b}_0$ for the mean square nonparametric estimation of $m(\cdot)$.

For $d = 2$, the optimal order of b_0^* is $n^{-1/5}$. Again this order goes to 0 faster than the order $n^{-1/6}$ of the optimal bandwidth for the nonparametric estimation of the regression function with two covariates.

However, for $d \geq 3$, we note that the order of b_0^* goes to 0 slowly than $\widehat{b}_0$. Hence our results show that optimal $\widehat{m}_n(\cdot)$ for estimating $f(\cdot)$ should use a very small bandwidth b_0. This suggests that $\widehat{m}_n(\cdot)$ should be less biased and should have a higher variance than the optimal Kernel regression estimator of the estimation setup. Such a finding parallels Wang, Cai, Brown and Levine (2008) who show that a similar result hold when estimating the conditional variance of a heteroscedastic regression error term. However Wang et *al.* (2008) do not give the order of the optimal bandwidth to be used for estimating the regression function in their heteroscedastic setup. These results show that estimators of $m(\cdot)$ with smaller bias should be preferred in our framework, compared to the case where the regression function $m(\cdot)$ is the parameter of interest.

4.3 Asymptotic normality

We give now an asymptotic normality of the estimator $\widehat{f}_{1n}(\epsilon)$.

Theorem 4.4. *Assume that*

$$(\mathbf{A_{11}}): \quad nb_0^{d+4} = O(1), \quad nb_0^4 b_1 = o(1), \quad nb_0^d b_1^3 \to \infty,$$

when n goes to ∞*. Then under* $(A_1) - (A_5)$*,* $(A_7) - (A_{10})$*, we have*

$$\sqrt{nb_1}\left(\widehat{f}_{1n}(\epsilon) - \overline{f}_{1n}(\epsilon)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(\epsilon)}{\mathbb{P}(X \in \mathcal{X}_0)} \int K_1^2(v)dv\right),$$

where

$$\overline{f}_{1n}(\epsilon) = f(\epsilon) + \frac{b_1^2}{2} f^{(2)}(\epsilon) \int v^2 K_1(v)dv + o\left(b_1^2\right).$$

The result of this theorem shows that the best choice b_1^* for the bandwidth b_1 should achieve the minimum of the Asymptotic Mean Integrated Square Error

$$\text{AMISE} = \frac{b_1^4}{4} \int \left(f^{(2)}(\epsilon)\right)^2 d\epsilon \left(\int v^2 K_1(v)dv\right)^2 + \frac{1}{nb_1\mathbb{P}(X \in \mathcal{X}_0)} \int K_1^2(v)dv,$$

leading to the optimal bandwidth

$$b_1^* = \left[\frac{\frac{1}{\mathbb{P}(X \in \mathcal{X}_0)} \displaystyle\int K_1^2(v)dv}{\displaystyle\int (f^{(2)}(\epsilon))^2 d\epsilon \left(\int v^2 K_1(v)dv \right)^2} \right]^{1/5} n^{-1/5}.$$

We also note that for $d \leq 2$, $b_1 = b_1^*$ and $b_0 = b_0^*$, Theorems 4.3 and 4.2 give

$$b_1 \asymp \left(\frac{1}{n}\right)^{\frac{1}{5}}, \quad b_0 \asymp \left(\frac{1}{n}\right)^{\frac{8}{5(2d+4)}},$$

which yields that

$$nb_0^{d+4} \asymp \left(\frac{1}{n}\right)^{\frac{12-2d}{5(2d+4)}}, \quad nb_0^4 b_1 \asymp \left(\frac{1}{n}\right)^{\frac{16-8d}{5(2d+4)}}, \quad nb_0^d b_1^3 \asymp \left(\frac{1}{n}\right)^{\frac{4d-8}{5(2d+4)}}.$$

This shows that for $d = 1$, the $(\mathbf{A_{11}})$ is realizable with the optimal bandwidths b_0^* and b_1^*. But with these bandwidths, the last constraint of $(\mathbf{A_{11}})$ is not satisfied for $d = 2$, since $nb_0^d b_1^3$ is bounded when $n \to \infty$.

5 Conclusion

The aim of this chapter was to study the nonparametric Kernel estimation of the probability density function of the regression error using the estimated residuals. The difference between the feasible estimator which uses the estimated residuals and the unfeasible one using the true residuals are studied. An optimal choice of the first-step bandwidth used to estimate the residuals is also proposed. Again, an asymptotic normality of the feasible Kernel estimator and its rate-optimality are established. One of the contributions of this paper is the analysis of the impact of the estimated residuals on the regression errors p.d.f. Kernel estimator.

In our setup, the strategy was to use an approach based on a two-steps procedure which, in a first step, replaces the unobserved residuals terms by some nonparametric estimators $\widehat{\varepsilon}_i$. In a second step, the "pseudo-observations"$\widehat{\varepsilon}_i$ are used to estimate the p.d.f $f(\cdot)$, as if they were the true ε_i's. If proceeding so can remedy the curse of dimensionality, a challenging issue was to measure the impact of the estimated residuals on the final estimator of $f(\cdot)$ in the first nonparametric step, and to find the order of the optimal first-step bandwidth b_0. For this choice of b_0, our results indicates that the optimal bandwidth to be used for estimating the regression function $m(\cdot)$ should be smaller than the optimal bandwidth for the mean square estimation of $m(\cdot)$. That is to say, the best estimator $\widehat{m}_n(\cdot)$ of the regression function $m(\cdot)$ needed for estimating $f(\cdot)$ should have a lower bias and a higher variance than the optimal Kernel regression of the

estimation setup. With this appropriate choice of b_0, it has been seen that for $d \leq 2$, the nonparametric estimator $\widehat{f}_{1n}(\epsilon)$ of f can reach the optimal rate $n^{-2/5}$, which corresponds to the exact consistency rate reached for the Kernel density estimator of real-valued variable. Hence our main conclusion is that for $d \leq 2$, the estimator $\widehat{f}_{1n}(\epsilon)$ used for estimating $f(\epsilon)$ is not affected by the curse of dimensionality, since there is no negative effect coming from the estimation of the residuals on the final estimator of $f(\epsilon)$.

6 Proofs section

Intermediate Lemmas for Proposition 2.1 and Theorem 4.1

Lemma 6.1. *Define, for $x \in \mathcal{X}_0$,*

$$\widehat{g}_n(x) = \frac{1}{nb_0^d} \sum_{i=1}^{n} K_0\left(\frac{X_i - x}{b_0}\right), \quad \overline{g}_n(x) = \mathbb{E}\left[\widehat{g}_n(x)\right].$$

Then under $(A_1) - (A_2)$, (A_4), (A_7) and (A_9), we have, when b_0 goes to 0,

$$\sup_{x \in \mathcal{X}_0} \left|\overline{g}_n(x) - g(x)\right| = O\left(b_0^2\right), \quad \sup_{x \in \mathcal{X}_0} \left|\widehat{g}_n(x) - \overline{g}_n(x)\right| = O_{\mathbb{P}}\left(b_0^4 + \frac{\ln n}{nb_0^d}\right)^{1/2},$$

and

$$\sup_{x \in \mathcal{X}_0} \left|\frac{1}{\widehat{g}_n(x)} - \frac{1}{g(x)}\right| = O_{\mathbb{P}}\left(b_0^4 + \frac{\ln n}{nb_0^d}\right)^{1/2}.$$

Lemma 6.2. *Under $(A_1) - (A_4)$, (A_7) and (A_9), we have*

$$\sup_{x \in \mathcal{X}_0} \left|\widehat{m}_n(x) - m(x)\right| = O_{\mathbb{P}}\left(b_0^4 + \frac{\ln n}{nb_0^d}\right)^{1/2}.$$

Lemma 6.3. *Define for $(x, y) \in \mathbb{R}^d \times \mathbb{R}$,*

$$f_n(\epsilon|x) = \frac{\frac{1}{nh_0^d h_1} \sum_{i=1}^{n} K_0\left(\frac{X_i - x}{h_0}\right) K_1\left(\frac{Y_i - m(x) - \epsilon}{h_1}\right)}{\frac{1}{nh_0^d} \sum_{i=1}^{n} K_0\left(\frac{X_i - x}{h_0}\right)},$$

Then under $(A_1) - (A_3)$, $(A_6) - (A_9)$, we have, when n goes to infinity,

$$\widetilde{f}_n(\epsilon|x) - f_n(\epsilon|x) = o_{\mathbb{P}}\left(\frac{1}{nh_0^d h_1}\right)^{1/2}.$$

Lemma 6.4. *Set, for $(x,y) \in \mathbb{R}^d \times \mathbb{R}$,*

$$\widetilde{\varphi}_{in}(x,y) = \frac{1}{h_0^d h_1} K_0\left(\frac{X_i - x}{h_0}\right) K_1\left(\frac{Y_i - y}{h_1}\right).$$

Then, under $(A_6) - (A_8)$, we have, for x in $\mathcal{X}_0$ and y in $\mathbb{R}$, h_0 and h_1 going to 0, and for some constant $C > 0$,

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\varphi}_{in}(x,y)\right] - \varphi(x,y) &= \frac{h_0^2}{2}\frac{\partial^2 \varphi(x,y)}{\partial^2 x}\int z K_0(z) z^\top dz + \frac{h_1^2}{2}\frac{\partial^2 \varphi(x,y)}{\partial^2 y}\int v^2 K_1(v) dv \\
&\quad + o\left(h_0^2 + h_1^2\right), \\
\operatorname{Var}\left[\widetilde{\varphi}_{in}(x,y)\right] &= \frac{\varphi(x,y)}{h_0^d h_1}\int\int K_0^2(z) K_1^2(v) dv dz + o\left(\frac{1}{h_0^d h_1}\right), \\
\mathbb{E}\left[\left|\widetilde{\varphi}_{in}(x,y) - \mathbb{E}\widetilde{\varphi}_{in}(x,y)\right|^3\right] &\leq \frac{C\varphi(x,y)}{h_0^{2d} h_1^2}\int\int \left|K_0(z) K_1(v)\right|^3 dz dv + o\left(\frac{1}{h_0^{2d} h_1^2}\right).
\end{aligned}$$

Lemma 6.5. *Set*

$$f_{in}(\epsilon) = \frac{\mathbb{1}\left(X_i \in \mathcal{X}_0\right)}{b_1 \mathbb{P}\left(X \in \mathcal{X}_0\right)} K_1\left(\frac{\varepsilon_i - \epsilon}{b_1}\right).$$

Then under (A_4), (A_5) and (A_8), we have, for b_1 going to 0, and for some constant $C > 0$,

$$\begin{aligned}
\mathbb{E} f_{in}(\epsilon) &= f(\epsilon) + \frac{b_1^2}{2} f^{(2)}(\epsilon)\int v^2 K_1(v) dv + o\left(b_1^2\right), \\
\operatorname{Var}\left(f_{in}(\epsilon)\right) &= \frac{f(\epsilon)}{b_1 \mathbb{P}\left(X \in \mathcal{X}_0\right)}\int K_1^2(v) dv + o\left(\frac{1}{b_1}\right), \\
\mathbb{E}\left|f_{in}(\epsilon) - \mathbb{E} f_{in}(\epsilon)\right|^3 &\leq \frac{C f(\epsilon)}{b_1^2 \mathbb{P}^2\left(X \in \mathcal{X}_0\right)}\int \left|K_1(v)\right|^3 dv + o\left(\frac{1}{b_1^2}\right).
\end{aligned}$$

Lemma 6.6. *Define*

$$\begin{aligned}
S_n &= \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right) K_1^{(1)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right), \\
T_n &= \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^2 K_1^{(2)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right), \\
R_n &= \sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^3 \int_0^1 (1-t)^2 K_1^{(3)}\left(\frac{\varepsilon_i - t(\widehat{m}_{in} - m(X_i)) - \epsilon}{b_1}\right) dt.
\end{aligned}$$

Then under $(A_1)-(A_5)$ and $(A_7)-(A_{10})$, we have, for b_0 and b_1 small enough,

$$\begin{aligned} S_n &= O_{\mathbb{P}}\left[b_0^2\left(nb_1^2+(nb_1)^{1/2}\right)+\left(nb_1^4+\frac{b_1}{b_0^d}\right)^{1/2}\right], \\ T_n &= O_{\mathbb{P}}\left[\left(nb_1^3+(nb_1)^{1/2}+\left(n^2b_0^db_1^3\right)^{1/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)\right], \\ R_n &= O_{\mathbb{P}}\left[\left(nb_1^3+\left(n^2b_0^db_1\right)^{1/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^{3/2}\right]. \end{aligned}$$

Lemma 6.7. *Under (A_5) and (A_8) we have, for some constant $C>0$, and for any ϵ in $\mathbb{R}$ and $p\in[0,2]$,*

$$\left|\int K_1^{(1)}\left(\frac{e-\epsilon}{b_1}\right)^2 e^p f(e)de\right|\le Cb_1, \qquad \left|\int K_1^{(1)}\left(\frac{e-\epsilon}{b_1}\right) e^p f(e)de\right|\le Cb_1^2, \tag{6.1}$$

$$\left|\int K_1^{(2)}\left(\frac{e-\epsilon}{b_1}\right)^2 e^p f(e)de\right|\le Cb_1, \qquad \left|\int K_1^{(2)}\left(\frac{e-\epsilon}{b_1}\right) e^p f(e)de\right|\le Cb_1^3, \tag{6.2}$$

$$\left|\int K_1^{(3)}\left(\frac{e-\epsilon}{b_1}\right)^2 e^p f(e)de\right|\le Cb_1, \qquad \left|\int K_1^{(3)}\left(\frac{e-\epsilon}{b_1}\right) e^p f(e)de\right|\le Cb_1^3. \tag{6.3}$$

Lemma 6.8. *Set*

$$\beta_{in}=\frac{\mathbb{1}\left(X_i\in\mathcal{X}_0\right)}{nb_0^d\widehat{g}_{in}}\sum_{j=1,j\neq i}^{n}\left(m(X_j)-m(X_i)\right)K_0\left(\frac{X_j-X_i}{b_0}\right).$$

Then, under $(A_1)-(A_5)$ and $(A_7)-(A_{10})$, we have, when b_0 and b_1 go to 0,

$$\sum_{i=1}^{n}\beta_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)=O_{\mathbb{P}}\left(b_0^2\right)\left(nb_1^2+(nb_1)^{1/2}\right).$$

Lemma 6.9. *Set*

$$\Sigma_{in}=\frac{\mathbb{1}\left(X_i\in\mathcal{X}_0\right)}{nb_0^d\widehat{g}_{in}}\sum_{j=1,j\neq i}^{n}\varepsilon_jK_0\left(\frac{X_j-X_i}{b_0}\right).$$

Then, under $(A_1)-(A_5)$ and $(A_7)-(A_{10})$, we have

$$\sum_{i=1}^{n}\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)=O_{\mathbb{P}}\left(nb_1^4+\frac{b_1}{b_0^d}\right)^{1/2}.$$

Lemma 6.10. *Let $\mathbb{E}_n[\cdot]$ be the conditional mean given $X_1, \ldots, X_n$. Then under $(A_1)-(A_5)$ and $(A_7)-(A_9)$, we have, for b_0 going to 0,*

$$\sup_{1\le i\le n} \mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^4\right] = O_{\mathbb{P}}\left(b_0^4 + \frac{1}{nb_0^d}\right)^2,$$
$$\sup_{1\le i\le n} \mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^6\right] = O_{\mathbb{P}}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3.$$

Lemma 6.11. *Assume that (A_4) and (A_7) hold. Then, for any $1 \le i \ne j \le n$, and for any ϵ in $\mathbb{R}$,*

$$(\widehat{m}_{in} - m(X_i), \varepsilon_i) \text{ and } (\widehat{m}_{jn} - m(X_j), \varepsilon_j)$$

are independent given $X_1, \ldots, X_n$, provided that $\|X_i - X_j\| \ge Cb_0$, for some constant $C > 0$.

Lemma 6.12. *Let $\mathrm{Var}_n(\cdot)$ and $Cov_n(\cdot)$ be respectively the conditional variance and the conditional covariance given $X_1, \ldots, X_n$, and set*

$$\zeta_{in} = \mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^2 K_1^{(2)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right).$$

Then under $(A_1) - (A_5)$ and $(A_7) - (A_9)$, we have, for n going to infinity,

$$\sum_{i=1}^{n} \mathrm{Var}_n\left(\zeta_{in}\right) = O_{\mathbb{P}}\left(nb_1\right)\left(b_0^4 + \frac{1}{nb_0^d}\right)^2,$$
$$\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\ne i}}^{n} \mathrm{Cov}_n\left(\zeta_{in}, \zeta_{jn}\right) = O_{\mathbb{P}}\left(n^2 b_0^d b_1^{7/2}\right)\left(b_0^4 + \frac{1}{nb_0^d}\right)^2.$$

All these lemmas are proved in Appendix A.

Proof of Proposition 2.1

Define $f_n(\epsilon|x)$ as in Lemma 6.3, and note that by this lemma, we have

$$\widetilde{f}_n(\epsilon|x) = f_n(\epsilon|x) + o_{\mathbb{P}}\left(\frac{1}{nh_0^d h_1}\right)^{1/2}. \tag{6.4}$$

The asymptotic distribution of the first term in (6.4) is derived by applying the Lyapounov Central Limit Theorem for triangular arrays (see e.g Billingsley 1968, Theorem 7.3). Define for $x \in \mathcal{X}_0$ and $y \in \mathbb{R}$,

$$\widetilde{\varphi}_n(x,y) = \frac{1}{nh_0^d h_1}\sum_{i=1}^{n} K_0\left(\frac{X_i - x}{h_0}\right) K_1\left(\frac{Y_i - y}{h_1}\right), \quad \widetilde{g}_n(x) = \frac{1}{nh_0^d}\sum_{i=1}^{n} K_0\left(\frac{X_i - x}{h_0}\right),$$

and observe that

$$f_n(\epsilon|x) = \frac{\widetilde{\varphi}_n\left(x, m(x)+\epsilon\right)}{\widetilde{g}_n(x)}\ . \tag{6.5}$$

Let now $\widetilde{\varphi}_{in}(x,y)$ be as in Lemma 6.4, and note that

$$\widetilde{\varphi}_n(x,y) = \frac{1}{n}\sum_{i=1}^{n}\Big(\widetilde{\varphi}_{in}(x,y) - \mathbb{E}\left[\widetilde{\varphi}_{in}(x,y)\right]\Big) + \mathbb{E}\left[\widetilde{\varphi}_{1n}(x,y)\right]. \tag{6.6}$$

The second and third inequalities in Lemma 6.4 give, since $h_0^d h_1$ goes to 0,

$$\frac{\sum_{i=1}^{n}\mathbb{E}\left|\widetilde{\varphi}_{in}(x,y) - \mathbb{E}\widetilde{\varphi}_{in}(x,y)\right|^3}{\left(\sum_{i=1}^{n}\operatorname{Var}\left[\widetilde{\varphi}_{in}(x,y)\right]\right)^3} \leq \frac{\frac{Cn\varphi(x,y)}{h_0^{2d}h_1^2}\int\int\left|K_0(z)K_1(v)\right|^3 dzdv + o\left(\frac{n}{h_0^{2d}h_1^2}\right)}{\left(\frac{n\varphi(x,y)}{h_0^d h_1}\int\int K_0^2(z)K_1^2(v)dvdz + o\left(\frac{n}{h_0^d h_1}\right)\right)^3} = O(h_0^d h_1) = o(1).$$

Hence the Lyapounov Central Limit Theorem gives, since $nh_0^d h_1$ diverges under ($\mathbf{A_0}$),

$$\frac{\sum_{i=1}^{n}\left\{\widetilde{\varphi}_{in}(x,y) - \mathbb{E}\left[\widetilde{\varphi}_{in}(x,y)\right]\right\}}{\left(\sum_{i=1}^{n}\operatorname{Var}\left[\widetilde{\varphi}_{in}(x,y)\right]\right)^{1/2}} \xrightarrow{d} \mathcal{N}(0,1),$$

so that

$$\frac{\sqrt{nh_0^d h_1}}{n}\sum_{i=1}^{n}\Big(\widetilde{\varphi}_{in}(x,y) - \mathbb{E}\left[\widetilde{\varphi}_{in}(x,y)\right]\Big) \xrightarrow{d} \mathcal{N}\left(0, \varphi(x,y)\int\int K_0^2(z)K_1^2(v)dzdv\right). \tag{6.7}$$

Further, a similar proof as the one of Lemma 6.1 gives

$$\frac{1}{\widetilde{g}_n(x)} = \frac{1}{g(x)} + O_{\mathbb{P}}\left(h_0^4 + \frac{\ln n}{nh_0^d}\right)^{1/2}. \tag{6.8}$$

Hence by this equality, it follows that, taking $y = m(x)+\epsilon$ in (6.7), and by (6.4)-(6.6),

$$\sqrt{nh_0^d h_1}\left(\widetilde{f}_n(\epsilon|x) - \overline{f}_n(\epsilon|x)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(\epsilon|x)}{g(x)}\int\int K_0^2(z)K_1^2(v)dzdv\right),$$

where

$$\overline{f}_n(\epsilon|x) = \frac{\mathbb{E}\left[\widetilde{\varphi}_{1n}\left(x, m(x)+\epsilon\right)\right]}{\widetilde{g}_n(x)}\ .$$

This yields the result of Proposition 2.1, since the first equality of Lemma 6.4 and (6.8) yield, for h_0 and h_1 small enough,

$$\begin{aligned}\overline{f}_n(\epsilon|x) \quad &= \quad f(\epsilon|x) + \frac{h_0^2}{2g(x)}\frac{\partial^2\varphi\left(x, m(x)+\epsilon\right)}{\partial^2 x}\int zK_0(z)z^\top dz \\ &\quad + \frac{h_1^2}{2g(x)}\frac{\partial^2\varphi(x, m(x)+\epsilon)}{\partial^2 y}\int v^2K_1(v)dv + o\left(h_0^2 + h_1^2\right).\square\end{aligned}$$

Proof of Theorem 4.1

The proof of the theorem is based upon the following equalities:

$$\begin{aligned}\widehat{f}_{1n}(\epsilon)-\widetilde{f}_{1n}(\epsilon) &= O_{\mathbb{P}}\left[b_0^2+\left(\frac{1}{n}+\frac{1}{n^2b_0^db_1^3}\right)^{1/2}\right]+O_{\mathbb{P}}\left[\frac{1}{(nb_1^5)^{1/2}}+\left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]\left(b_0^4+\frac{1}{nb_0^d}\right)\\ &\quad+O_{\mathbb{P}}\left[\frac{1}{b_1}+\left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]\left(b_0^4+\frac{1}{nb_0^d}\right)^{3/2}, \qquad (6.9)\end{aligned}$$

and

$$\widetilde{f}_{1n}(\epsilon)-f(\epsilon)=O_{\mathbb{P}}\left(b_1^4+\frac{1}{nb_1}\right)^{1/2}. \qquad (6.10)$$

Indeed, since $\widehat{f}_{1n}(\epsilon)-f(\epsilon)=\left(\widetilde{f}_{1n}(\epsilon)-f(\epsilon)\right)+\widehat{f}_{1n}(\epsilon)-\widetilde{f}_{1n}(\epsilon)$, it then follows by (6.10) and (6.9) that

$$\begin{aligned}\widehat{f}_{1n}(\epsilon)-f(\epsilon) &= O_{\mathbb{P}}\left[b_1^4+\frac{1}{nb_1}+b_0^4+\frac{1}{n}+\frac{1}{n^2b_0^db_1^3}+\left(\frac{1}{(nb_1^5)^{1/2}}+\left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right)^2\left(b_0^4+\frac{1}{nb_0^d}\right)^2\right]^{1/2}\\ &\quad+O_{\mathbb{P}}\left[\left(\frac{1}{b_1}+\left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right)^2\left(b_0^4+\frac{1}{nb_0^d}\right)^3\right]^{1/2}.\end{aligned}$$

This yields the result of the Theorem, since under (A_9) and (A_{10}), we have

$$\frac{1}{n}=O\left(\frac{1}{nb_1}\right),\quad \frac{1}{n^2b_0^db_1^3}=O\left(\frac{b_0^d}{b_1^3}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2.$$

Hence, it remains to prove (6.9) and (6.10). For this, define S_n, R_n and T_n as in Lemma 6.6. Since $\widehat{\varepsilon}_i-\varepsilon_i=-\left(\widehat{m}_{in}-m(X_i)\right)$ and that K_1 is three times continuously differentiable under (A_8), the third-order Taylor expansion with integral remainder gives

$$\begin{aligned}\widehat{f}_{1n}(\epsilon)-\widetilde{f}_{1n}(\epsilon) &= \frac{1}{b_1\sum_{i=1}^n \mathbb{1}(X_i\in\mathcal{X}_0)}\sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left[K_1\left(\frac{\widehat{\varepsilon}_i-\epsilon}{b_1}\right)-K_1\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\\ &= -\frac{1}{b_1\sum_{i=1}^n \mathbb{1}(X_i\in\mathcal{X}_0)}\left(\frac{S_n}{b_1}-\frac{T_n}{2b_1^2}+\frac{R_n}{2b_1^3}\right).\end{aligned}$$

Therefore, since

$$\sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)=n\left(\mathbb{P}\left(X\in\mathcal{X}_0\right)+o_{\mathbb{P}}(1)\right),$$

by the Law of large numbers, Lemma 6.6 then gives

$$
\begin{aligned}
\widehat{f}_{1n}(\epsilon)-\widetilde{f}_{1n}(\epsilon) &= O_{\mathbb{P}}\left(\frac{1}{nb_1^2}\right)S_n+O_{\mathbb{P}}\left(\frac{1}{nb_1^3}\right)T_n+O_{\mathbb{P}}\left(\frac{1}{nb_1^4}\right)R_n \\
&= O_{\mathbb{P}}\left[b_0^2\left(1+\frac{1}{(nb_1^3)^{1/2}}\right)+\left(\frac{1}{n}+\frac{1}{n^2b_0^db_1^3}\right)^{1/2}\right] \\
&\quad +O_{\mathbb{P}}\left[1+\frac{1}{(nb_1^5)^{1/2}}+\left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]\left(b_0^4+\frac{1}{nb_0^d}\right)+O_{\mathbb{P}}\left[\frac{1}{b_1}+\left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]\left(b_0^4+\frac{1}{nb_0^d}\right)^{3/2}.
\end{aligned}
$$

This yields (6.9), since under (A_9) and (A_{10}), we have $b_0\to 0$, $nb_0^{d+2}\to\infty$ and $nb_1^3\to\infty$, so that

$$
\begin{aligned}
b_0^2\left(1+\frac{1}{(nb_1^3)^{1/2}}\right) &\asymp O\left(b_0^2\right), \quad \left(b_0^4+\frac{1}{nb_0^d}\right)=O\left(b_0^2\right), \\
\left[1+\frac{1}{(nb_1^5)^{1/2}}+\left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]\left(b_0^4+\frac{1}{nb_0^d}\right) &= O\left(b_0^2\right)+\left[\frac{1}{(nb_1^5)^{1/2}}+\left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]\left(b_0^4+\frac{1}{nb_0^d}\right).
\end{aligned}
$$

For (6.10), note that

$$
\mathbb{E}_n\left[\left(\widetilde{f}_{1n}(\epsilon)-f(\epsilon)\right)^2\right]=\mathrm{Var}_n\left(\widetilde{f}_{1n}(\epsilon)\right)+\left(\mathbb{E}_n\left[\widetilde{f}_{1n}(\epsilon)\right]-f(\epsilon)\right)^2, \tag{6.11}
$$

with, using (A_4),

$$
\mathrm{Var}_n\left(\widetilde{f}_{1n}(\epsilon)\right)=\frac{1}{\left(b_1\sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)\right)^2}\sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)\mathrm{Var}\left[K_1\left(\frac{\varepsilon-\epsilon}{b_1}\right)\right].
$$

Therefore, since the Cauchy-Schwarz inequality gives

$$
\mathrm{Var}\left[K_1\left(\frac{\varepsilon-\epsilon}{b_1}\right)\right]\le\mathbb{E}\left[K_1^2\left(\frac{\varepsilon-\epsilon}{b_1}\right)\right]\le b_1\int K_1^2(v)f(\epsilon+b_1v)dv,
$$

this bound and the equality above yield, under (A_5) and (A_8),

$$
\mathrm{Var}_n\left(\widetilde{f}_{1n}(\epsilon)\right)\le\frac{C}{b_1\sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)}=O_{\mathbb{P}}\left(\frac{1}{nb_1}\right). \tag{6.12}
$$

For the second term in (6.11), we have

$$
\mathbb{E}_n\left[\widetilde{f}_{1n}(\epsilon)\right]=\frac{1}{b_1\sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)}\sum_{i=1}^n \mathbb{1}\left(X_i\in\mathcal{X}_0\right)\mathbb{E}\left[K_1\left(\frac{\varepsilon-\epsilon}{b_1}\right)\right]. \tag{6.13}
$$

By (A_8), K_1 is symmetric, has a compact support, with $\int vK_1(v)=0$ and $\int K_1(v)dv=1$. Therefore, since under (A_5) f has bounded continuous second order derivatives, this yields for some $\theta=\theta(\epsilon,b_1v)$,

$$
\begin{aligned}
\mathbb{E}\left[K_1\left(\frac{\varepsilon-\epsilon}{b_1}\right)\right] &= b_1\int K_1(v)f(\epsilon+b_1v)dv \\
&= b_1\int K_1(v)\left[f(\epsilon)+b_1vf^{(1)}(\epsilon)+\frac{b_1^2v^2}{2}f^{(2)}(\epsilon+\theta b_1v)\right]dv \\
&= b_1f(\epsilon)+\frac{b_1^3}{2}\int v^2K_1(v)f^{(2)}(\epsilon+\theta b_1v)dv.
\end{aligned}
$$

Hence this equality and (6.13) give

$$\mathbb{E}_n\left[\widetilde{f}_{1n}(\epsilon)\right] = f(\epsilon) + \frac{b_1^2}{2}\int v^2 K_1(v) f^{(2)}(\epsilon + \theta b_1 v)dv,$$

so that

$$\left(\mathbb{E}_n\left[\widetilde{f}_{1n}(\epsilon)\right] - f(\epsilon)\right)^2 = O_{\mathbb{P}}\left(b_1^4\right).$$

Combining this result with (6.12) and (6.11), we obtain, by the Tchebychev inequality,

$$\widetilde{f}_{1n}(\epsilon) - f(\epsilon) = O_{\mathbb{P}}\left(b_1^4 + \frac{1}{nb_1}\right)^{1/2}.$$

This proves (6.10), and then achieves the proof of the theorem. □

Proof of Theorem 4.2

Recall that

$$R_n(b_0, b_1) = b_0^4 + \left[\frac{1}{(nb_1^5)^{1/2}} + \left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]^2\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 + \left[\frac{1}{b_1} + \left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]^2\left(b_0^4 + \frac{1}{nb_0^d}\right)^3,$$

and note that

$$\left(\frac{1}{n^2 b_1^3}\right)^{\frac{1}{d+4}} = \max\left\{\left(\frac{1}{n^2 b_1^3}\right)^{\frac{1}{d+4}}, \left(\frac{1}{n^3 b_1^7}\right)^{\frac{1}{2d+4}}\right\}$$

if and only if $n^{4-d} b_1^{d+16} \to \infty$. To find the order of b_0^*, we shall deal with the cases $nb_0^{d+4} \to \infty$ and $nb_0^{d+4} = O(1)$.

First assume that $nb_0^{d+4} \to \infty$. More precisely, we suppose that b_0 is in $\left[(u_n/n)^{1/(d+4)}, +\infty\right)$, where $u_n \to \infty$. Since $1/(nb_0^d) = O(b_0^4)$ for all these b_0, we have

$$\left(b_0^4 + \frac{1}{nb_0^d}\right)^2 \asymp \left(b_0^4\right)^2, \quad \left(b_0^4 + \frac{1}{nb_0^d}\right)^3 \asymp \left(b_0^4\right)^3.$$

Hence the order of b_0^* is computed by minimizing the function

$$b_0 \to b_0^4 + \left[\frac{1}{(nb_1^5)^{1/2}} + \left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]^2\left(b_0^4\right)^2 + \left[\frac{1}{b_1} + \left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]^2\left(b_0^4\right)^3.$$

Since this function is increasing with b_0, the minimum of $R_n(\cdot, b_1)$ is achieved for $b_{0*} = (u_n/n)^{1/(d+4)}$. We shall prove later on that this choice of b_{0*} is irrelevant compared to the one arising when $nb_0^{d+4} = O(1)$.

Consider now the case $nb_0^{d+4} = O(1)$ i.e $b_0^4 = O\left(1/(nb_0^d)\right)$. This gives

$$\begin{aligned}\left[\frac{1}{(nb_1^5)^{1/2}} + \left(\frac{b_0^d}{b_1^3}\right)^{1/2}\right]^2 \left(b_0^4 + \frac{1}{nb_0^d}\right)^2 &\asymp \left(\frac{1}{nb_1^5} + \frac{b_0^d}{b_1^3}\right)\left(\frac{1}{n^2 b_0^{2d}}\right), \\ \left[\frac{1}{b_1} + \left(\frac{b_0^d}{b_1^7}\right)^{1/2}\right]^2 \left(b_0^4 + \frac{1}{nb_0^d}\right)^3 &\asymp \left(\frac{1}{b_1^2} + \frac{b_0^d}{b_1^7}\right)\left(\frac{1}{n^3 b_0^{3d}}\right).\end{aligned}$$

Moreover if $nb_0^d b_1^4 \to \infty$, we have, since $nb_0^{2d} \to \infty$ under (A_9),

$$\left(\frac{1}{nb_1^5} + \frac{b_0^d}{b_1^3}\right)\left(\frac{1}{n^2 b_0^{2d}}\right) \asymp \frac{b_0^d}{b_1^3}\left(\frac{1}{n^2 b_0^{2d}}\right), \quad \left(\frac{1}{b_1^2} + \frac{b_0^d}{b_1^7}\right)\left(\frac{1}{n^3 b_0^{3d}}\right) = O\left(\frac{b_0^d}{b_1^3}\right)\left(\frac{1}{n^2 b_0^{2d}}\right).$$

Hence the order of b_0^* is obtained by finding the minimum of the function $b_0^4 + \left(1/n^2 b_0^d b_1^3\right)$. The minimization of this function gives a solution b_0 such that

$$b_0 \asymp \left(\frac{1}{n^2 b_1^3}\right)^{\frac{1}{d+4}}, \quad R_n(b_0, b_1) \asymp \left(\frac{1}{n^2 b_1^3}\right)^{\frac{4}{d+4}}.$$

This value satisfies the constraints $nb_0^{d+4} = O(1)$ and $nb_0^d b_1^4 \to \infty$ when $n^{4-d} b_1^{d+16} \to \infty$.

If now $nb_0^{d+4} = O(1)$ but $nb_0^d b_1^4 = O(1)$, we have, since $nb_0^{2d} \to \infty$,

$$\frac{1}{nb_1^5}\left(\frac{1}{n^2 b_0^{2d}}\right) = O\left(\frac{b_0^d}{b_1^7}\right)\left(\frac{1}{n^3 b_0^{3d}}\right), \quad \frac{1}{b_1^2}\left(\frac{1}{n^3 b_0^{3d}}\right) = O\left(\frac{b_0^d}{b_1^3}\right)\left(\frac{1}{n^2 b_0^{2d}}\right) = O\left(\frac{b_0^d}{b_1^7}\right)\left(\frac{1}{n^3 b_0^{3d}}\right).$$

In this case, b_0^* is obtained by minimizing the function $b_0^4 + \left(1/n^3 b_0^{2d} b_1^7\right)$, for which the solution b_0 verifies

$$b_0 \asymp \left(\frac{1}{n^3 b_1^7}\right)^{\frac{1}{2d+4}}, \quad R_n(b_0, b_1) \asymp \left(\frac{1}{n^3 b_1^7}\right)^{\frac{4}{2d+4}}.$$

This solution fulfills the constraint $nb_0^d b_1^4 = O(1)$ when $n^{4-d} b_1^{d+16} = O(1)$. Hence we can conclude that for $b_0^4 = O\left(1/(nb_0^d)\right)$, the bandwidth b_0^* satisfies

$$b_0^* \asymp \max\left\{\left(\frac{1}{n^2 b_1^3}\right)^{\frac{1}{d+4}}, \left(\frac{1}{n^3 b_1^7}\right)^{\frac{1}{2d+4}}\right\},$$

which leads to

$$R_n\left(b_0^*, b_1\right) \asymp \max\left\{\left(\frac{1}{n^2 b_1^3}\right)^{\frac{4}{d+4}}, \left(\frac{1}{n^3 b_1^7}\right)^{\frac{4}{2d+4}}\right\}.$$

We need now to compare the solution b_0^* to the candidate $b_{0*} = (u_n/n)^{1/(d+4)}$ obtained when $nb_0^{d+4} \to \infty$. For this, we must do a comparison between the orders of $R_n(b_0^*, b_1)$ and $R_n(b_{0*}, b_1)$. Since $R_n(b_0, b_1) \geq b_0^4$, we have $R_n(b_{0*}, b_1) \geq (u_n/n)^{4/(d+4)}$, so that, for n large enough,

$$\begin{aligned}\frac{R_n(b_0^*, b_1)}{R_n(b_{0*}, b_1)} &\leq C\left[\left(\frac{1}{n^2 b_1^3}\right)^{\frac{1}{d+4}} + \left(\frac{1}{n^3 b_1^7}\right)^{\frac{4}{2d+4}}\right]\left(\frac{n}{u_n}\right)^{\frac{4}{d+4}} \\ &= o(1) + O\left(\frac{1}{u_n}\right)^{\frac{4}{d+4}}\left(\frac{1}{nb_1^{\frac{7(d+4)}{d+8}}}\right)^{\frac{4(d+8)}{(2d+4)(d+4)}} = o(1),\end{aligned}$$

using $u_n \to \infty$ and that $n^{(d+8)} b_1^{7(d+4)} \to \infty$ by (A_{10}). This shows that $R_n(b_0^*, b_1) \leq R_n(b_{0*}, b_1)$ for n large enough. Hence the Theorem is proved, since b_0^* is the best candidate for the minimization of $R_n(\cdot, b_1)$. $\square$

Proof of Theorem 4.3

Recall that Theorem 4.2 gives

$$AMSE(b_1) + R_n(b_0^*, b_1) \asymp r_1(b_1) + r_2(b_1) + r_3(b_1) = F(b_1),$$

where

$$\begin{aligned}
r_1(h) &= h^4 + \frac{1}{nh}, \quad \arg\min r_1(h) \asymp n^{-1/5} = h_1^*, \quad \min r_1(h) \asymp (h_1^*)^4 = n^{-4/5}, \\
r_2(h) &= h^4 + \frac{1}{n^{\frac{8}{d+4}} h^{\frac{12}{d+4}}}, \quad \arg\min r_2(h) \asymp n^{-\frac{2}{d+7}} = h_2^*, \quad \min r_3(h) \asymp (h_2^*)^4 = n^{-\frac{8}{d+7}}, \\
r_3(h) &= h^4 + \frac{1}{n^{\frac{12}{2d+4}} h^{\frac{28}{2d+4}}}, \quad \arg\min r_3(h) \asymp n^{-\frac{3}{2d+11}} = h_3^*, \quad \min r_3(h) \asymp (h_3^*)^4 = n^{-\frac{12}{2d+11}}.
\end{aligned}$$

Each $r_j(h)$ decreases on $[0, \arg\min r_j(h)]$ and increases on $(\arg\min r_j(h), \infty)$ and that $r_j(h) \asymp h^4$ on $(\arg\min r_j(h), \infty)$. Moreover $\min r_2(h) = o(r_3(h))$ and $h_2^* = o(h_3^*)$ for all possible dimension d, so that $\min\{r_2(h) + r_3(h)\} \asymp (h_3^*)^4 = n^{-\frac{12}{2d+11}}$ and $\arg\min\{r_2(h) + r_3(h)\} \asymp h_3^* = n^{-\frac{3}{2d+11}}$.

Observe now that $\min\{r_2(h)+r_3(h)\} = O(\min r_1(h))$ is equivalent to $n^{-\frac{12}{2d+11}} = O\left(n^{-4/5}\right)$ which holds if and only if $d \leq 2$. Hence assume that $d \leq 2$. Since $n^{-\frac{12}{2d+11}} = O\left(n^{-4/5}\right)$ also gives $\arg\min\{r_2(h)+r_3(h)\} \asymp h_3^* = O(h_1^*)$, we have

$$\min F(b_1) \asymp n^{-4/5} \text{ and } \arg\min F(b_1) \asymp n^{-1/5}.$$

The case $d > 2$ is symmetric with

$$\min F(b_1) \asymp n^{-\frac{12}{2d+11}} \text{ and } \arg\min F(b_1) \asymp n^{-\frac{3}{2d+11}}.$$

This ends the proof of the Theorem. $\square$

Proof of Theorem 4.4

Observe that the Tchebychev inequality gives

$$\sum_{i=1}^{n} \mathbb{1}(X_i \in \mathcal{X}_0) = n\mathbb{P}(X \in \mathcal{X}_0) \left[1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)\right],$$

so that

$$\widetilde{f}_{1n}(\epsilon) = \left[1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)\right] f_n(\epsilon),$$

where

$$f_n(\epsilon) = \frac{1}{nb_1 \mathbb{P}(X \in \mathcal{X}_0)} \sum_{i=1}^{n} \mathbb{1}(X_i \in \mathcal{X}_0) K_1 \left(\frac{\varepsilon_i - \epsilon}{b_1}\right).$$

Therefore

$$\widehat{f}_{1n}(\epsilon) - \mathbb{E}f_n(\epsilon) = (f_n(\epsilon) - \mathbb{E}f_n(\epsilon)) + \left(\widehat{f}_{1n}(\epsilon) - \widetilde{f}_{1n}(\epsilon)\right) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) f_n(\epsilon). \tag{6.14}$$

Let now $f_{in}(\epsilon)$ be as in Lemma 6.5, and note that $f_n(\epsilon) = (1/n)\sum_{i=1}^{n} f_{in}(\epsilon)$. The second and the third claims in Lemma 6.5 yield, since b_1 goes to 0 under (A_{10}),

$$\frac{\sum_{i=1}^{n} \mathbb{E}\left|f_{in}(\epsilon) - \mathbb{E}f_{in}(\epsilon)\right|^3}{\left(\sum_{i=1}^{n} \mathrm{Var} f_{in}(\epsilon)\right)^3} \le \frac{\frac{Cnf(\epsilon)}{\mathbb{P}(X \in \mathcal{X}_0)^2 b_1^2} \int |K_1(v)|^3 dv + o\left(\frac{n}{b_1^2}\right)}{\left(\frac{nf(\epsilon)}{\mathbb{P}(X \in \mathcal{X}_0) b_1} \int K_1^2(v) dv + o\left(\frac{n}{b_1}\right)\right)^3} = O(b_1) = o(1).$$

Hence the Lyapounov Central Limit Theorem gives, since nb_1 diverges under (A_{10}),

$$\frac{f_n(\epsilon) - \mathbb{E}f_n(\epsilon)}{\sqrt{\mathrm{Var} f_n(\epsilon)}} = \frac{f_n(\epsilon) - \mathbb{E}f_n(\epsilon)}{\sqrt{\frac{\mathrm{Var} f_{in}(\epsilon)}{n}}} \xrightarrow{d} \mathcal{N}(0, 1),$$

which yields, using the second equality in Lemma 6.5,

$$\sqrt{nb_1}\left(f_n(\epsilon) - \mathbb{E}f_n(\epsilon)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(\epsilon)}{\mathbb{P}(X \in \mathcal{X}_0)} \int K_1^2(v) dv\right). \tag{6.15}$$

Moreover, note that for $nb_0^d b_1^3 \to \infty$ and $nb_0^{2d} \to \infty$,

$$\frac{1}{nb_1^5}\left(\frac{1}{nb_0^d}\right)^2 + \left(\frac{1}{b_1^2} + \frac{b_0^d}{b_1^7}\right)^2 \left(\frac{1}{nb_0^d}\right)^3 = O\left(\frac{1}{n^2 b_0^d b_1^3}\right).$$

Therefore, since by Assumptions (A_{11}) and (A_9), we have $b_0^4 = O\left(1/(nb_0^d)\right)$, $nb_0^d b_1^3 \to \infty$ and that $nb_0^{2d} \to \infty$, the equality above and (6.9) then give

$$\begin{aligned}
\widehat{f}_{1n}(\epsilon) - \widetilde{f}_{1n}(\epsilon) &\asymp O_{\mathbb{P}}\left[b_0^4 + \frac{1}{n} + \frac{1}{n^2 b_0^d b_1^3} + \left(\frac{1}{nb_1^5} + \frac{b_0^d}{b_1^3}\right)\left(\frac{1}{nb_0^d}\right)^2 + \left(\frac{1}{b_1^2} + \frac{b_0^d}{b_1^7}\right)\left(\frac{1}{nb_0^d}\right)^3\right]^{1/2} \\
&\asymp O_{\mathbb{P}}\left(b_0^4 + \frac{1}{n} + \frac{1}{n^2 b_0^d b_1^3}\right)^{1/2}.
\end{aligned}$$

Hence for b_1 going to 0, we have

$$\sqrt{nb_1}\left(\widehat{f}_{1n}(\epsilon) - \widetilde{f}_{1n}(\epsilon)\right) = O_{\mathbb{P}}\left[nb_1\left(b_0^4 + \frac{1}{n} + \frac{1}{n^2 b_0^d b_1^3}\right)\right]^{1/2} = o_{\mathbb{P}}(1),$$

since $nb_0^4 b_1 = o(1)$ and that $nb_0^d b_1^2 \to \infty$ under Assumption (A_{11}). Combining the above result with (6.15) and (6.14), we obtain

$$\sqrt{nb_1}\left(\widehat{f}_{1n}(\epsilon) - \mathbb{E}f_n(\epsilon)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(\epsilon)}{\mathbb{P}(X \in \mathcal{X}_0)} \int K_1^2(v) dv\right).$$

This ends the proof the Theorem, since the first result of Lemma 6.5 gives

$$\mathbb{E}f_n(\epsilon) = \mathbb{E}f_{1n}(\epsilon) = f(\epsilon) + \frac{b_1^2}{2} f^{(2)}(\epsilon) \int v^2 K_1(v) dv + o\left(b_1^2\right) := \overline{f}_{1n}(\epsilon). \square$$

Appendix A: Proof of the intermediate results

Proof of Lemma 6.1

First note that by (A_7), we have $\int z K_0(z) dz = 0$ and $\int K_0(z) dz = 1$. Therefore, since K_0 is continuous and has a compact support, (A_1), (A_2) and a second-order Taylor expansion, yield, for b_0 small enough and any x in $\mathcal{X}_0$,

$$\begin{aligned}
|\overline{g}_n(x) - g(x)| &= \left| \frac{1}{b_0^d} \int K_0\left(\frac{z-x}{b_0}\right) g(z) dz - g(x) \right| = \left| \int K_0(z) \left[g(x + b_0 z) - g(x)\right] dz \right| \\
&= \left| \int K_0(z) \left[b_0 g^{(1)}(x) z + \frac{b_0^2}{2} z g^{(2)}(x + \theta b_0 z) z^\top \right] dz \right|, \ \theta = \theta(x, b_0 z) \in [0,1] \\
&= \left| b_0 g^{(1)}(x) \int z K_0(z) dz + \frac{b_0^2}{2} \int z g^{(2)}(x + \theta b_0 z) z^\top K_0(z) dz \right| \\
&= \frac{b_0^2}{2} \left| \int z g^{(2)}(x + \theta b_0 z) z^\top K_0(z) dz \right| \leq C b_0^2,
\end{aligned}$$

so that

$$\sup_{x \in \mathcal{X}_0} |\overline{g}_n(x) - g(x)| = O\left(b_0^2\right).$$

This gives the first equality of the lemma. To prove the two last equalities in the Lemma, note that it is sufficient to show that

$$\sup_{x \in \mathcal{X}_0} |\widehat{g}_n(x) - \overline{g}_n(x)| = O_{\mathbb{P}} \left(\frac{\ln n}{n b_0^d}\right)^{1/2},$$

since $\bar{g}_n(x)$ is asymptotically bounded away from 0 over $\mathcal{X}_0$ and that $|\overline{g}_n(x) - g(x)| = O(b_0^2)$ uniformly for x in $\mathcal{X}_0$. This follows from Theorem 1 in Einmahl and Mason (2005). $\square$

Proof of Lemma 6.2

For the first equality in the lemma, set

$$\widehat{r}_n(x) = \frac{1}{n b_0^d} \sum_{j=1}^n Y_j K_0\left(\frac{X_j - x}{b_0}\right), \quad \overline{r}_n(x) = \mathbb{E}\left[\widehat{r}_n(x)\right],$$

and observe that

$$\sup_{x \in \mathcal{X}_0} |\widehat{m}_n(x) - m(x)| \leq \sup_{x \in \mathcal{X}_0} \left| \widehat{m}_n(x) - \frac{\overline{r}_n(x)}{\overline{g}_n(x)} \right| + \sup_{x \in \mathcal{X}_0} \frac{1}{|\overline{g}_n(x)|} \left| \overline{r}_n(x) - \overline{g}_n(x) m(x) \right|. \tag{A.1}$$

Consider the first term of (A.1). Note that $\mathbb{E}^{1/4}\left[Y^4|X=x\right] \leq |m(x)| + \mathbb{E}^{1/4}\left[\varepsilon^4\right]$. The compactness of $\mathcal{X}$ from (A_1), the continuity of $m(\cdot)$ from (A_3) and (A_4) then give that $\mathbb{E}\left[Y^4|X=x\right] < \infty$ uniformly for $x \in \mathcal{X}_0$. Hence under (A_9), Theorem 2 in Einmahl and Mason (2005) gives

$$\sup_{x\in\mathcal{X}_0}\left|\widehat{m}_n(x) - \frac{\overline{r}_n(x)}{\overline{g}_n(x)}\right| = O_{\mathbb{P}}\left(\frac{\ln n}{nb_0^d}\right)^{1/2}.$$

For the second term in (A.1), a second-order Taylor expansion gives, as in the proof of Lemma 6.1,

$$\sup_{x\in\mathcal{X}_0}\left|\overline{r}_n(x) - \overline{g}_n(x)m(x)\right| = O(b_0^2).$$

This gives the result of lemma since Lemma 6.1 implies that $\overline{g}_n(x)$ is bounded away from 0 over $\mathcal{X}_0$ uniformly in x and for b_0 small enough. □

Proof of Lemma 6.3

Note that under (A_8), the Taylor expansion with integral remainder gives, for any $x \in \mathcal{X}_0$ and any integer $i \in [1,n]$,

$$K_1\left(\frac{Y_i - \widehat{m}_n(x) - \epsilon}{h_1}\right) = K_1\left(\frac{Y_i - m(x) - \epsilon}{h_1}\right) - \frac{1}{h_1}\left(\widehat{m}_n(x) - m(x)\right)\int_0^1 K_1^{(1)}\left(\frac{Y_i - \theta_n(x,t)}{h_1}\right)dt,$$

where $\theta_n(x,t) = m(x) + \epsilon + t\left(\widehat{m}_n(x) - m(x)\right)$. Therefore

$$\widetilde{f}_n(\epsilon|x) = f_n(\epsilon|x) - \frac{\widehat{m}_n(x) - m(x)}{\widetilde{g}_n(x)}\left[\frac{1}{nh_0^dh_1^2}\sum_{i=1}^n K_0\left(\frac{X_i - x}{h_0}\right)\int_0^1 K_1^{(1)}\left(\frac{Y_i - \theta_n(x,t)}{h_1}\right)dt\right]. \tag{A.2}$$

Now, observe that if $X_i = z$ and $y \in \mathbb{R}$, the change of variable $e = y - m(z) + h_1v$ gives, under $(A_1) - (A_5)$ and (A_7),

$$\begin{aligned}
\mathbb{E}_n\left|K_1^{(1)}\left(\frac{Y_i - y}{h_1}\right)\right| &= \mathbb{E}\left|K_1^{(1)}\left(\frac{\varepsilon_i + m(z) - y}{h_1}\right)\right| \\
&= \int\left|K_1^{(1)}\left(\frac{e + m(z) - y}{h_1}\right)\right| f(e)de \\
&= h_1\int |K_1^{(1)}(v)| f\left((y - m(z) + h_1v)\right)dv \leq Ch_1.
\end{aligned}$$

Hence

$$\sup_{1\leq i\leq n}\int_0^1 \mathbb{E}_n\left|K_1^{(1)}\left(\frac{Y_i - \theta_n(x,t)}{h_1}\right)\right|dt \leq Ch_1.$$

With the help of this result and Lemma 6.1, we have

$$\begin{aligned}
&\mathbb{E}_n\left|\frac{1}{nh_0^dh_1}\sum_{i=1}^n K_0\left(\frac{X_i-x}{h_0}\right)\int_0^1 K_1^{(1)}\left(\frac{Y_i-\theta_n(x,t)}{h_1}\right)dt\right|\\
&\leq \frac{1}{nh_0^dh_1}\sum_{i=1}^n\left|K_0\left(\frac{X_i-x}{h_0}\right)\right|\times\sup_{1\leq i\leq n}\int_0^1\mathbb{E}_n\left|K_1^{(1)}\left(\frac{Y_i-\theta_n(x,t)}{h_1}\right)\right|dt\\
&\leq \frac{C}{nh_0^d}\sum_{i=1}^n\left|K_0\left(\frac{X_i-x}{h_0}\right)\right| = O_{\mathbb{P}}(1),
\end{aligned}$$

so that

$$\frac{1}{nh_0^dh_1^2}\sum_{i=1}^n K_0\left(\frac{X_i-x}{h_0}\right)\int_0^1 K_1^{(1)}\left(\frac{Y_i-\theta_n(x,t)}{h_1}\right)dt = O_{\mathbb{P}}\left(\frac{1}{h_1}\right).$$

Hence from (A.2), (6.8), Lemma 6.2 and Assumption ($\mathbf{A_0}$), we deduce

$$\widetilde{f}_n(\epsilon|x) = f_n(\epsilon|x) + O_{\mathbb{P}}\left(\frac{1}{h_1}\right)\left(b_0^4+\frac{\ln n}{nb_0^d}\right)^{1/2} = f_n(\epsilon|x) + o\left(\frac{1}{nh_0^dh_1}\right)^{1/2}.\Box$$

Proof of Lemma 6.4 and Lemma 6.5

We just give the proof of Lemma 6.4, the proof of Lemma 6.5 being very similar. For the first equality of Lemma 6.4, note that

$$\begin{aligned}
\mathbb{E}\left[\widetilde{\varphi}_{in}(x,y)\right] &= \frac{1}{h_0^dh_1}\int\int K_0\left(\frac{x_1-x}{h_0}\right)K_1\left(\frac{y_1-y}{h_1}\right)\varphi(x_1,y_1)dx_1dy_1\\
&= \int\int K_0(z)K_1(v)\varphi\left(x+h_0z,y+h_1v\right)dzdv.
\end{aligned}$$

A second-order Taylor expansion gives under (A_6), for z in the support of K_0, v in the support of K_1, and h_0, h_1 small enough,

$$\begin{aligned}
&\varphi\left(x+h_0z,y+h_1v\right)-\varphi(x,y)\\
&= h_0\frac{\partial\varphi(x,y)}{\partial x}z^\top + h_1\frac{\partial\varphi(x,y)}{\partial y}v\\
&\quad+\frac{h_0^2}{2}z\frac{\partial^2\varphi(x+\theta h_0z,y+\theta h_1v)}{\partial^2x}z^\top + h_1h_0v\frac{\partial^2\varphi(x+\theta h_0z,y+\theta h_1v)}{\partial x\partial y}z^\top\\
&\quad+\frac{h_1^2}{2}\frac{\partial^2\varphi(x+\theta h_0z,y+\theta h_1v)}{\partial^2y}v^2,
\end{aligned}$$

for some $\theta = \theta(x, y, h_0 z, h_1 v)$ in $[0,1]$. This gives, since $\int K_0(z)dz = \int K_1(v)dv = 1$, $\int zK_0(z)dz$ and $\int vK_1(v)dv$ vanish under $(A_7) - (A_8)$, and by the Lebesgue Dominated Convergence Theorem,

$$\begin{aligned}
&\mathbb{E}\left[\widetilde{\varphi}_{in}(x,y)\right] - \varphi(x,y) - \frac{h_0^2}{2}\frac{\partial^2\varphi(x,y)}{\partial^2 x}\int zK_0(z)z^\top dz - \frac{h_1^2}{2}\frac{\partial^2\varphi(x,y)}{\partial^2 y}\int v^2K_1(v)dv \\
= \ & \frac{h_0^2}{2}\int\int z\left(\frac{\partial^2\varphi(x+\theta h_0 z, y+\theta h_1 v)}{\partial^2 x} - \frac{\partial^2\varphi(x,y)}{\partial^2 x}\right) z^\top K_0(z)K_1(v)dzdv \\
&+h_1h_0\int\int v\left(\frac{\partial^2\varphi(x+\theta h_0 z, y+\theta h_1 v)}{\partial x\partial y} - \frac{\partial^2\varphi(x,y)}{\partial x\partial y}\right) z^\top K_0(z)K_1(v)dzdv \\
&+\frac{h_1^2}{2}\int\int \left(\frac{\partial^2\varphi(x+\theta h_0 z, y+\theta h_1 v)}{\partial^2 y} - \frac{\partial^2\varphi(x,y)}{\partial^2 y}\right) v^2 K_0(z)K_1(v)dzdv \\
= \ & o(h_0^2 + h_1^2).
\end{aligned}$$

This proves the first equality of Lemma 6.4. The second equality in Lemma follows similarly, since

$$\begin{aligned}
&\mathrm{Var}[\widetilde{\varphi}_{in}(x,y)] = \mathbb{E}\left[\widetilde{\varphi}_{in}^2(x,y)\right] - \left(\mathbb{E}\left[\widetilde{\varphi}_{in}(x,y)\right]\right)^2 \\
= \ & \frac{1}{h_0^d h_1}\int\int \varphi\left(x+h_0 z, y+h_1 v\right) K_0^2(z)K_1^2(v)dzdv + O(1) \\
= \ & \frac{\varphi(x,y)}{h_0^d h_1}\int\int K_0^2(z)K_1^2(v)dzdv + o\left(\frac{1}{h_0^d h_1}\right).
\end{aligned}$$

The last statement of Lemma 6.4 is immediate, since the Triangular and Convex inequalities give

$$\begin{aligned}
\mathbb{E}\left|\widetilde{\varphi}_{in}(x,y) - \mathbb{E}\widetilde{\varphi}_{in}(x,y)\right|^3 &\leq C\mathbb{E}\left|\widetilde{\varphi}_{in}(x,y)\right|^3 \\
&\leq \frac{C\varphi(x,y)}{h_0^{2d}h_1^2}\int\int \left|K_0(z)K_1(v)\right|^3 dzdv + o\left(\frac{1}{h_0^{2d}h_1^2}\right). \square
\end{aligned}$$

Proof of Lemma 6.6

The order of S_n follows from Lemma 6.8 and Lemma 6.9. In fact, since

$$\begin{aligned}
\mathbb{1}(X_i \in \mathcal{X}_0)\left(\widehat{m}_{in} - m(X_i)\right) &= \frac{\mathbb{1}(X_i \in \mathcal{X}_0)}{nb_0^d\widehat{g}_{in}}\sum_{j=1, j\neq i}^{n}\left(m(X_j) + \varepsilon_j - m(X_i)\right)K_0\left(\frac{X_j - X_i}{b_0}\right) \\
&= \beta_{in} + \Sigma_{in},
\end{aligned}$$

Lemma 6.8 and Lemma 6.9 give

$$S_n = O_{\mathbb{P}}\left[b_0^2\left(nb_1^2 + (nb_1)^{1/2}\right) + \left(nb_1^4 + \frac{b_1}{b_0^d}\right)^{1/2}\right],$$

which gives the result for S_n.

For T_n, define for any $1 \leq i \leq n$,

$$\mathbb{E}_{in}[\cdot] = \mathbb{E}_n\left[X_1, \ldots, X_n, \varepsilon_k, k \neq i\right].$$

Therefore, since $(\widehat{m}_{in} - m(X_i))$ depends only upon $(X_1, \ldots, X_n, \varepsilon_k, k \neq i)$, we have

$$\begin{aligned}
\mathbb{E}_n[T_n] &= \mathbb{E}_n\left[\sum_{i=1}^n \mathbb{E}_{in}\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)(\widehat{m}_{in} - m(X_i))^2 K_1^{(2)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right]\right] \\
&= \mathbb{E}_n\left[\sum_{i=1}^n \mathbb{1}\left(X_i \in \mathcal{X}_0\right)(\widehat{m}_{in} - m(X_i))^2 \mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right]\right],
\end{aligned}$$

with, using (A_4) and Lemma 6.7-(6.2),

$$\left|\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right]\right| = \left|\int K_1^{(2)}\left(\frac{e - \epsilon}{b_1}\right) f(e) de\right| \leq C b_1^3.$$

Hence this bound, the equality above, the Cauchy-Schwarz inequality and Lemma 6.10 yield that

$$\begin{aligned}
\left|\mathbb{E}_n\left[T_n\right]\right| &\leq C b_1^3 \sum_{i=1}^n \mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)(\widehat{m}_{in} - m(X_i))^2\right] \\
&\leq C n b_1^3 \left(\sup_{1 \leq i \leq n} \mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)(\widehat{m}_{in} - m(X_i))^4\right]\right)^{1/2} \\
&\leq O_{\mathbb{P}}\left(n b_1^3\right)\left(b_0^4 + \frac{1}{n b_0^d}\right). \qquad \text{(A.3)}
\end{aligned}$$

For the conditional variance of T_n, Lemma 6.12 gives

$$\begin{aligned}
\mathrm{Var}_n(T_n) &= \sum_{i=1}^n \mathrm{Var}_n\left(\zeta_{in}\right) + \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \mathrm{Cov}_n\left(\zeta_{in}, \zeta_{jn}\right) \\
&= O_{\mathbb{P}}\left(n b_1\right)\left(b_0^4 + \frac{b_1}{n b_0^d}\right)^2 + O_{\mathbb{P}}\left(n^2 b_0^d b_1^{7/2}\right)\left(b_0^4 + \frac{1}{n b_0^d}\right)^2.
\end{aligned}$$

Therefore, since b_1 goes to 0 under (A_{10}), this order and (A.3) yield, applying the Tchebychev inequality,

$$\begin{aligned}
T_n &= O_{\mathbb{P}}\left[\left(n b_1^3\right)\left(b_0^4 + \frac{1}{n b_0^d}\right) + (n b_1)^{1/2}\left(b_0^4 + \frac{b_1}{n b_0^d}\right) + \left(n^2 b_0^d b_1^{7/2}\right)^{1/2}\left(b_0^4 + \frac{1}{n b_0^d}\right)\right] \\
&= O_{\mathbb{P}}\left[\left(n b_1^3 + (n b_1)^{1/2} + \left(n^2 b_0^d b_1^3\right)^{1/2}\right)\left(b_0^4 + \frac{1}{n b_0^d}\right)\right].
\end{aligned}$$

which gives the result for T_n.

We now compute the order of R_n. For this, define

$$\begin{aligned}
I_{in} &= \int_0^1 (1-t)^2 K_1^{(3)}\left(\frac{\varepsilon_i - t(\widehat{m}_{in} - m(X_i)) - \epsilon}{b_1}\right) dt, \\
R_{in} &= \mathbb{1}\left(X_i \in \mathcal{X}_0\right)(\widehat{m}_{in} - m(X_i))^3 I_{in},
\end{aligned}$$

and note that $R_n = \sum_{i=1}^n R_{in}$. The order of R_n is derived by computing its conditional mean and its

conditional variance. For the conditional mean, observe that

$$\begin{aligned}\mathbb{E}_n[R_n] &= \mathbb{E}_n\left[\sum_{i=1}^{n}\mathbb{E}_{in}\left[R_{in}\right]\right]\\ &= \mathbb{E}_n\left[\sum_{i=1}^{n}\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^3\mathbb{E}_{in}\left[I_{in}\right]\right],\end{aligned}$$

with, using (A_4) and Lemma 6.7-(6.3),

$$\begin{aligned}\left|\mathbb{E}_{in}\left[I_{in}\right]\right| &= \left|\int_0^1(1-t)^2\left[\int K_1^{(3)}\left(\frac{e-t(\widehat{m}_{in}-m(X_i))-\epsilon}{b_1}\right)f(e)de\right]dt\right|\\ &\leq Cb_1^3.\end{aligned}$$

Therefore the Holder inequality and Lemma 6.10 yield

$$\begin{aligned}\left|\mathbb{E}_n\left[R_n\right]\right| &\leq Cb_1^3\sum_{i=1}^{n}\mathbb{E}_n\left[\left|\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)\right|^3\right]\\ &\leq Cb_1^3\sum_{i=1}^{n}\mathbb{E}_n^{3/4}\left[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^4\right]\\ &\leq O_{\mathbb{P}}\left(nb_1^3\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^{3/2}. \qquad (A.4)\end{aligned}$$

For the conditional covariance of R_n, note that Lemma 6.11 allows to write

$$\operatorname{Var}_n\left(R_n\right)=\sum_{i=1}^{n}\operatorname{Var}_n\left(R_{in}\right)+\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\Big(\|X_i-X_j\|\leq Cb_0\Big)\operatorname{Cov}_n\left(R_{in},R_{jn}\right), \qquad (A.5)$$

and consider the first term in (A.5). We have

$$\operatorname{Var}_n\left(R_{in}\right)\leq\mathbb{E}_n\left[R_{in}^2\right]\leq\mathbb{E}_n\Big[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^6\mathbb{E}_{in}\left[I_{in}^2\right]\Big],$$

with, using (A_4), the Cauchy-Schwarz inequality and Lemma 6.7-(6.3),

$$\begin{aligned}\mathbb{E}_{in}\left[I_{in}^2\right] &\leq C\mathbb{E}_{in}\left[\int_0^1 K_1^{(3)}\left(\frac{\varepsilon_i-t(\widehat{m}_{in}-m(X_i))-\epsilon}{b_1}\right)^2dt\right]\\ &\leq C\int_0^1\left[\int K_1^{(3)}\left(\frac{e-t(\widehat{m}_{in}-m(X_i))-\epsilon}{b_1}\right)^2f(e)de\right]dt\\ &\leq Cb_1,\end{aligned}$$

so that

$$\operatorname{Var}_n\left(R_{in}\right)\leq Cb_1\mathbb{E}_n\left[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^6\right].$$

Therefore form Lemma 6.10, we deduce

$$\begin{aligned}\sum_{i=1}^{n}\mathrm{Var}_n\left(R_{in}\right) &\leq Cnb_1\sup_{1\leq i\leq n}\mathbb{E}_n\left[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^6\right]\\ &\leq O_{\mathbb{P}}\left(nb_1\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^3. \end{aligned}\tag{A.6}$$

For the second term in (A.5), the Cauchy-Schwarz inequality gives, with the help of the above result for $\mathrm{Var}_n\left(R_{in}\right)$,

$$\begin{aligned}\left|\mathrm{Cov}_n\left(R_{in},R_{jn}\right)\right| &\leq \left(\mathrm{Var}_n\left(R_{in}\right)\mathrm{Var}_n\left(R_{jn}\right)\right)^{1/2}\\ &\leq Cb_1\sup_{1\leq i\leq n}\mathbb{E}_n\left[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^6\right].\end{aligned}$$

Hence by Lemma 6.10 and the Markov inequality, we have

$$\begin{aligned}&\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\Big(\|X_i-X_j\|\leq Cb_0\Big)\left|\mathrm{Cov}_n\left(R_{in},R_{jn}\right)\right|\\ &\leq O_{\mathbb{P}}\left(b_1\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^3\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\Big(\|X_i-X_j\|\leq Cb_0\Big)\\ &\leq O_{\mathbb{P}}\left(b_1\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^3\left(n^2b_0^d\right).\end{aligned}$$

This order, (A.6) and (A.5) give, since nb_0^d diverges under (A_9),

$$\mathrm{Var}\left(R_n\right)=O_{\mathbb{P}}\left(b_0^4+\frac{1}{nb_0^d}\right)^3\left(n^2b_0^db_1\right).$$

Finally, with the help of this result, (A.4) and the Tchebychev inequality, we arrive at

$$\begin{aligned}R_n &= O_{\mathbb{P}}\left[\left(nb_1^3\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^{3/2}+\left(n^2b_0^db_1\right)^{1/2}\left(b_0^4+\frac{1}{nb_0^d}\right)^{3/2}\right]\\ &= O_{\mathbb{P}}\left[\left(nb_1^3+\left(n^2b_0^db_1\right)^{1/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^{3/2}\right].\square\end{aligned}$$

Proof of Lemma 6.7

Set $h_p(e)=e^pf(e)$, $p\in[0,2]$. For the first inequality of (6.1), note that under (A_5) and (A_8), the change of variable $e=\epsilon+b_1v$ give, for any integer $\ell\in[1,3]$,

$$\begin{aligned}\left|\int K_1^{(\ell)}\left(\frac{e-\epsilon}{b_1}\right)^2e^pf(e)de\right| &= \left|b_1\int K_1^{(\ell)}(v)^2h_p(\epsilon+b_1v)dv\right|\\ &\leq b_1\sup_{t\in\mathbb{R}}|h_p(t)|\int|K_1^{(\ell)}(v)^2|dv\\ &\leq Cb_1,\end{aligned}\tag{A.7}$$

which yields the first inequality in (6.1). For the second inequality in (6.1), observe that $f(\cdot)$ has a bounded continuous derivative under (A_5), and that $\int K_1^{(\ell)}(v)dv = 0$ under (A_8). Therefore, since $h_p(\cdot)$ has bounded second order derivatives under (A_7), the Taylor inequality yields that

$$\begin{aligned}\left|\int K_1^{(\ell)}\left(\frac{e-\epsilon}{b_1}\right)e^p f(e)de\right| &= b_1\left|\int K_1^{(\ell)}(v)\left[h_p(\epsilon+b_1v)-h_p(\epsilon)\right]\right|dv \\ &\leq b_1^2\sup_{t\in\mathbb{R}}|h_p^{(1)}(t)|\int|vK_1^{(\ell)}(v)|dv \leq Cb_1^2.\end{aligned}$$

which completes the proof of (6.1).

The first inequalities of (6.2) and (6.3) follow directly from (A.7). The second bounds in (6.2) and (6.3) are proved simultaneously. For this, note that for any integer $\ell \in \{2,3\}$,

$$\int K_1^{(\ell)}\left(\frac{e-\epsilon}{b_1}\right)h_p(e)de = b_1\int K_1^{(\ell)}(v)h_p(\epsilon+b_1v)dv.$$

Under (A_8), $K_1(\cdot)$ is symmetric, has a compact support and two continuous derivatives, with $\int K_1^{(\ell)}(v)dv = 0$ and $\int vK_1^{(\ell)}(v)dv = 0$ for $\ell \in \{2,3\}$. Hence, since by (A_5) h_p has bounded continuous second order derivatives, this gives for some $\theta = \theta(\epsilon, b_1v)$,

$$\begin{aligned}\left|\int K_1^{(\ell)}\left(\frac{e-\epsilon}{b_1}\right)h_p(e)de\right| &= \left|b_1\int K_1^{(\ell)}(v)\left[h_p(\epsilon+b_1v)-h_p(\epsilon)\right]dv\right| \\ &= \left|b_1\int K_1^{(\ell)}(v)\left[b_1vh_p^{(1)}(\epsilon)+\frac{b_1^2v^2}{2}h_p^{(2)}(\epsilon+\theta b_1v)\right]dv\right| \\ &= \left|\frac{b_1^3}{2}\int v^2K_1^{(\ell)}(v)h_p^{(2)}(\epsilon+\theta b_1v)dv\right| \\ &\leq \frac{b_1^3}{2}\sup_{t\in\mathbb{R}}|h_p^{(2)}(t)|\int\left|v^2K_1^{(\ell)}(v)\right|dv \leq Cb_1^3.\square\end{aligned}$$

Proof of Lemma 6.8

Assumption (A_4) and Lemma 6.7-(6.1) give

$$\begin{aligned}\left|\mathbb{E}_n\left[\sum_{i=1}^n\beta_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right| &= \left|\mathbb{E}\left[K_1^{(1)}\left(\frac{\varepsilon-\epsilon}{b_1}\right)\right]\sum_{i=1}^n\beta_{in}\right| \leq Cnb_1^2\max_{1\leq i\leq n}|\beta_{in}|, \\ \mathrm{Var}_n\left[\sum_{i=1}^n\beta_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right] &\leq \sum_{i=1}^n\beta_{in}^2\mathbb{E}\left[K_1^{(1)}\left(\frac{\varepsilon-\epsilon}{b_1}\right)^2\right] \leq Cnb_1\max_{1\leq i\leq n}|\beta_{in}|^2.\end{aligned}$$

Hence the (conditional) Markov inequality gives

$$\sum_{i=1}^n\beta_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right) = O_{\mathbb{P}}\left(nb_1^2+(nb_1)^{1/2}\right)\max_{1\leq i\leq n}|\beta_{in}|,$$

so that the lemma follows if we can prove that

$$\sup_{1\le i\le n} |\beta_{in}| = O_{\mathbb{P}}\left(b_0^2\right), \tag{A.8}$$

as established now. For this, define

$$\zeta_j(x) = \mathbb{1}\left(x\in\mathcal{X}_0\right)\left(m(X_j)-m(x)\right)K_0\left(\frac{X_j-x}{b_0}\right),\quad \nu_{in}(x) = \frac{1}{(n-1)b_0^d}\sum_{j=1,j\neq i}^{n}\left(\zeta_j(x)-\mathbb{E}[\zeta_j(x)]\right),$$

and $\bar{\nu}_n(x)=\mathbb{E}[\zeta_j(x)]/b_0^d$, so that

$$\beta_{in} = \frac{n-1}{n}\frac{\nu_{in}(X_i)+\bar{\nu}_n(X_i)}{\widehat{g}_{in}}.$$

For $\max_{1\le i\le n}|\bar{\nu}_n(X_i)|$, first observe that a second-order Taylor expansion applied successively to $g(\cdot)$ and $m(\cdot)$ give, for b_0 small enough, and for any x, z in $\mathcal{X}$,

$$\begin{aligned}
&[m(x+b_0z)-m(x)]\,g(x+b_0z)\\
&= \left[b_0 m^{(1)}(x)z+\frac{b_0^2}{2}zm^{(2)}(x+\zeta_1 b_0 z)z^\top\right]\left[g(x)+b_0g^{(1)}(x)z+\frac{b_0^2}{2}zg^{(2)}(x+\zeta_2 b_0 z)z^\top\right],
\end{aligned}$$

for some $\zeta_1=\zeta_1(x,b_0z)$ and $\zeta_2=\zeta_2(x,b_0z)$ in $[0,1]$. Therefore, since $\int zK(z)dz=0$ under (A_7), it follows that, by (A_1), (A_2) and (A_3),

$$\begin{aligned}
\max_{1\le i\le n}|\bar{\nu}_n(X_i)| &\le \sup_{x\in\mathcal{X}_0}|\bar{\nu}_n(x)| = \sup_{x\in\mathcal{X}_0}\left|\int\left(m(x+b_0z)-m(x)\right)K_0(z)g(x+b_0z)dz\right|\\
&\le Cb_0^2. \qquad\qquad (A.9)
\end{aligned}$$

Consider now the term $\max_{1\le i\le n}|\nu_{in}(X_i)|$. The Bernstein inequality (see e.g. Serfling (2002)) and (A_4) give, for any $t>0$,

$$\begin{aligned}
\mathbb{P}\left(\max_{1\le i\le n}|\nu_{in}(X_i)|\ge t\right) &\le \sum_{i=1}^n \mathbb{P}\left(|\nu_{in}(X_i)|\ge t\right)\le \sum_{i=1}^n\int \mathbb{P}\left(|\nu_{in}(x)|\ge t\,|X_i=x\right)g(x)dx\\
&\le 2n\exp\left(-\frac{(n-1)t^2}{2\sup_{x\in\mathcal{X}_0}\mathrm{Var}(\zeta_j(x)/b_0^d)+\frac{4M}{3b_0^d}t}\right),
\end{aligned}$$

where M is such that $\sup_{x\in\mathcal{X}_0}|\zeta_j(x)|\le M$. The definition of $\mathcal{X}_0$ given in (A_2), (A_3), (A_7) and the standard Taylor expansion yield, for b_0 small enough,

$$\sup_{x\in\mathcal{X}_0}|\zeta_j(x)|\le Cb_0,\quad \sup_{x\in\mathcal{X}_0}\mathrm{Var}(\zeta_j(x)/b_0^d)\le\frac{1}{b_0^d}\sup_{x\in\mathcal{X}_0}\int\left(m(x+b_0z)-m(x)\right)^2K_0^2(z)g(x+b_0z)dz\le\frac{Cb_0^2}{b_0^d},$$

so that, for any $t\ge0$,

$$\mathbb{P}\left(\max_{1\le i\le n}|\nu_{in}(X_i)|\ge t\right)\le 2n\exp\left(-\frac{(n-1)b_0^dt^2/b_0^2}{C+Ct/b_0}\right).$$

This gives

$$\mathbb{P}\left(\max_{1\le i\le n}|\nu_{in}(X_i)|\ge\left(\frac{b_0^2\ln n}{(n-1)b_0^d}\right)^{1/2}t\right)\le 2n\exp\left(-\frac{t^2\ln n}{C+Ct\left(\frac{\ln n}{(n-1)b_0^d}\right)^{1/2}}\right)=o(1),$$

provided that t is large enough and under (A_9). It then follows that

$$\max_{1\le i\le n}|\nu_{in}(X_i)|=O_{\mathbb{P}}\left(\frac{b_0^2\ln n}{nb_0^d}\right)^{1/2}.$$

This bound, (A.9) and Lemma 6.1 show that (A.8) is proved, since $b_0^2\ln n/(nb_0^d)=O\left(b_0^4\right)$ under (A_9), and that

$$\beta_{in}=\frac{n-1}{n}\frac{\nu_{in}(X_i)+\bar{\nu}_n(X_i)}{\widehat{g}_{in}}.\square$$

Proof of Lemma 6.9

Note that (A_4) gives that Σ_{in} is independent of ε_i, and that $\mathbb{E}_n[\Sigma_{in}]=0$. This yields

$$\mathbb{E}_n\left[\sum_{i=1}^n\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]=0. \tag{A.10}$$

Moreover, observe that

$$\begin{aligned}&\mathrm{Var}_n\left[\sum_{i=1}^n\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\\ =\;&\sum_{i=1}^n\mathrm{Var}_n\left[\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]+\sum_{i=1}^n\sum_{\substack{j=1\\ j\ne i}}^n\mathrm{Cov}_n\left[\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right),\Sigma_{jn}K_1^{(1)}\left(\frac{\varepsilon_j-\epsilon}{b_1}\right)\right].\end{aligned} \tag{A.11}$$

For the sum of variances in (A.11), Lemma 6.7-(6.1) and (A_4) give

$$\begin{aligned}\sum_{i=1}^n\mathrm{Var}_n\left[\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]&\le\sum_{i=1}^n\mathbb{E}_n\left[\Sigma_{in}^2\right]\mathbb{E}\left[K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)^2\right]\\ &\le\frac{Cb_1\sigma^2}{(nb_0^d)^2}\sum_{i=1}^n\sum_{\substack{j=1\\ j\ne i}}^n\frac{\mathbb{1}(X_i\in\mathcal{X}_0)}{\widehat{g}_{in}^2}K_0^2\left(\frac{X_j-X_i}{b_0}\right)\\ &\le\frac{Cb_1\sigma^2}{nb_0^d}\sum_{i=1}^n\frac{\mathbb{1}(X_i\in\mathcal{X}_0)\widetilde{g}_{in}}{\widehat{g}_{in}^2},\end{aligned} \tag{A.12}$$

where $\sigma^2=\mathrm{Var}(\varepsilon)$ and

$$\widetilde{g}_{in}=\frac{1}{nb_0^d}\sum_{j=1,j\ne i}^nK_0^2\left(\frac{X_j-X_i}{b_0}\right).$$

For the sum of conditional covariances in (A.11), observe that by (A_4) we have

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathrm{Cov}_n\left[\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right),\Sigma_{jn}K_1^{(1)}\left(\frac{\varepsilon_j-\epsilon}{b_1}\right)\right]\\
&\quad=\quad\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathbb{E}_n\left[\Sigma_{in}\Sigma_{jn}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)K_1^{(1)}\left(\frac{\varepsilon_j-\epsilon}{b_1}\right)\right]\\
&\quad=\quad\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\frac{\mathbb{1}(X_i\in\mathcal{X}_0)\mathbb{1}(X_j\in\mathcal{X}_0)}{(nb_0^d)^2\widehat{g}_{in}\widehat{g}_{jn}}\sum_{\substack{k=1\\k\neq i}}^{n}\sum_{\substack{\ell=1\\\ell\neq j}}^{n}K_0\left(\frac{X_k-X_i}{b_0}\right)K_0\left(\frac{X_\ell-X_j}{b_0}\right)\mathbb{E}\left[\xi_{ki}\xi_{\ell j}\right],
\end{aligned}$$

where

$$\xi_{ki}=\varepsilon_kK_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right).$$

Moreover, under (A_4), it is seen that for $k\neq\ell$, $\mathbb{E}[\xi_{ki}\xi_{\ell j}]=0$ when $\mathrm{Card}\{i,j,k,\ell\}\geq 3$. Therefore the symmetry of K_0 yields that

$$\begin{aligned}
&\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathrm{Cov}_n\left[\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right),\Sigma_{jn}K_1^{(1)}\left(\frac{\varepsilon_j-\epsilon}{b_1}\right)\right]\\
&\quad=\quad\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\frac{\mathbb{1}(X_i\in\mathcal{X}_0)\mathbb{1}(X_j\in\mathcal{X}_0)}{(nb_0^d)^2\widehat{g}_{in}\widehat{g}_{jn}}K_0^2\left(\frac{X_j-X_i}{b_0}\right)\mathbb{E}^2\left[\varepsilon K_1^{(1)}\left(\frac{\varepsilon-\epsilon}{b_1}\right)\right]\\
&\qquad+\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\frac{\mathbb{1}(X_i\in\mathcal{X}_0)\mathbb{1}(X_j\in\mathcal{X}_0)}{(nb_0^d)^2\widehat{g}_{in}\widehat{g}_{jn}}\sum_{\substack{k=1\\k\neq i,j}}^{n}K_0\left(\frac{X_k-X_i}{b_0}\right)K_0\left(\frac{X_k-X_j}{b_0}\right)\mathbb{E}[\varepsilon^2]\mathbb{E}^2\left[K_1^{(1)}\left(\frac{\varepsilon-\epsilon}{b_1}\right)\right].
\end{aligned}$$

Therefore, since

$$\sup_{1\leq j\leq n}\left(\frac{\mathbb{1}\left(X_j\in\mathcal{X}_0\right)}{|\widehat{g}_{jn}|}\right)=O_{\mathbb{P}}(1)$$

by Lemma 6.1, Lemma 6.7-(6.1) and (A_4) then give

$$\begin{aligned}
&\left|\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathrm{Cov}_n\left[\Sigma_{in}K_1^{(1)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right),\Sigma_{jn}K_1^{(1)}\left(\frac{\varepsilon_j-\epsilon}{b_1}\right)\right]\right|\\
&\quad=\quad O_{\mathbb{P}}\left(\frac{b_1^4}{nb_0^d}\right)\sum_{i=1}^{n}\frac{\mathbb{1}(X_i\in\mathcal{X}_0)\widetilde{g}_{in}}{|\widehat{g}_{in}|}+O_{\mathbb{P}}(b_1^4)\sum_{i=1}^{n}\frac{\mathbb{1}(X_i\in\mathcal{X}_0)|g_{in}|}{|\widehat{g}_{in}|}\ ,
\end{aligned}\tag{A.13}$$

where $\widetilde{g}_{in}$ is defined as in (A.12) and

$$g_{in}=\frac{1}{(nb_0^d)^2}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{\substack{k=1\\k\neq j,i}}^{n}K_0\left(\frac{X_k-X_i}{b_0}\right)K_0\left(\frac{X_k-X_j}{b_0}\right).$$

The order of the first term in (A.13) follows from Lemma 6.1, which gives

$$\sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)\widetilde{g}_{in}}{|\widehat{g}_{in}|} = O_{\mathbb{P}}(n). \tag{A.14}$$

Again, by Lemma 6.1, we have

$$\sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)|g_{in}|}{|\widehat{g}_{in}|} = O_{\mathbb{P}}(1) \sum_{i=1}^{n} \mathbb{1}\left(X_i \in \mathcal{X}_0\right)|g_{in}|,$$

with, using the changes of variables $x_1 = x_3 + b_0 z_1$, $x_2 = x_3 + b_0 z_2$,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i=1}^{n} \mathbb{1}\left(X_i \in \mathcal{X}_0\right)|g_{in}|\right] &\leq \frac{Cn^3}{(nb_0^d)^2} \mathbb{E}\left|K_0\left(\frac{X_3 - X_1}{b_0}\right) K_0\left(\frac{X_3 - X_2}{b_0}\right)\right| \\
&\leq \frac{Cn^3 b_0^{2d}}{(nb_0^d)^2} \int\int\int |K_0(z_1)K_0(z_2)|\, g(x_3 + b_0 z_1) g(x_3 + b_0 z_2) g(x_3) dz_1 dz_2 dx_3.
\end{aligned}$$

These bounds and the equality above, give under (A_2) and (A_7),

$$\sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)|g_{in}|}{|\widehat{g}_{in}|} = O_{\mathbb{P}}(n).$$

Hence from (A.14), (A.13), (A.12), (A.11) and Lemma 6.1, we deduce, for b_1 small enough,

$$\begin{aligned}
&\mathrm{Var}_n\left[\sum_{i=1}^{n} \Sigma_{in} K_1^{(1)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)\right] \\
&= O_{\mathbb{P}}\left(\frac{b_1}{nb_0^d}\right) \sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)\widetilde{g}_{in}}{\widehat{g}_{in}^2} + O_{\mathbb{P}}\left(\frac{b_1^4}{nb_0^d}\right) \sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)\widetilde{g}_{in}}{|\widehat{g}_{in}|} + O_{\mathbb{P}}(b_1^4) \sum_{i=1}^{n} \frac{\mathbb{1}(X_i \in \mathcal{X}_0)|g_{in}|}{|\widehat{g}_{in}|} \\
&= O_{\mathbb{P}}\left(\frac{b_1}{b_0^d} + \frac{b_1^4}{b_0^d} + nb_1^4\right) = O_{\mathbb{P}}\left(\frac{b_1}{b_0^d} + nb_1^4\right).
\end{aligned}$$

Finally, this order, (A.10) and the Tchebychev inequality give

$$\sum_{i=1}^{n} \Sigma_{in} K_1^{(1)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right) = O_{\mathbb{P}}\left(\frac{b_1}{b_0^d} + nb_1^4\right)^{1/2}. \square$$

Proof of Lemma 6.10

Define β_{in} as in Lemma 6.8 and set

$$g_{in} = \frac{1}{nb_0^d} \sum_{j=1, j\neq i}^{n} K_0^4\left(\frac{X_j - X_i}{b_0}\right), \quad \widetilde{g}_{in} = \frac{1}{nb_0^d} \sum_{j=1, j\neq i}^{n} K_0^2\left(\frac{X_j - X_i}{b_0}\right).$$

The proof of the lemma is based on the following bound:

$$\mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^k\right] \leq C\left[\beta_{in}^k + \frac{\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\widetilde{g}_{in}^{k/2}}{(nb_0^d)^{(k/2)}\widehat{g}_{in}^k}\right], \quad k \in \{4, 6\}. \tag{A.15}$$

Indeed, taking successively $k = 4$ and $k = 6$ in (A.15), we have, by (A.8), Lemma 6.1 and (A_9),

$$\begin{aligned}\sup_{1\le i\le n} \mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^4\right] &= O_{\mathbb{P}}\left(b_0^8 + \frac{1}{(nb_0^d)^2}\right) = O_{\mathbb{P}}\left(b_0^4 + \frac{1}{nb_0^d}\right)^2,\\ \sup_{1\le i\le n} \mathbb{E}_n\left[\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\left(\widehat{m}_{in} - m(X_i)\right)^6\right] &= O_{\mathbb{P}}\left(b_0^{12} + \frac{1}{(nb_0^d)^3}\right) = O_{\mathbb{P}}\left(b_0^4 + \frac{1}{nb_0^d}\right)^3,\end{aligned}$$

which gives the results of the Lemma. Hence it remains to prove (A.15). For this, define β_{in} and Σ_{in} respectively as in Lemma 6.8 and Lemma 6.9. Since $\mathbb{1}(X_i \in \mathcal{X}_0)\left(\widehat{m}_{in} - m(X_i)\right) = \beta_{in} + \Sigma_{in}$, and that β_{in} depends only on $(X_1, \ldots, X_n)$, this gives, for $k \in \{4, 6\}$

$$\mathbb{E}_n\left[\mathbb{1}(X_i \in \mathcal{X}_0)\left(\widehat{m}_{in} - m(X_i)\right)^k\right] \le C\beta_{in}^k + C\mathbb{E}_n\left[\Sigma_{in}^k\right]. \tag{A.16}$$

The order of the second term of bound (A.16) is computed by applying Theorem 2 in Whittle (1960) or the Marcinkiewicz-Zygmund inequality (see e.g Chow and Teicher, 2003, p. 386). These inequalities show that for linear form $L = \sum_{j=1}^n a_j\zeta_j$ with independent mean-zero random variables $\zeta_1, \ldots, \zeta_n$, it holds that, for any $k \ge 1$,

$$\mathbb{E}\left|L^k\right| \le C(k)\left[\sum_{j=1}^n a_j^2 \mathbb{E}^{2/k}\left|\zeta_j^k\right|\right]^{k/2},$$

where $C(k)$ is a positive real depending only on k. Now, observe that for any $i \in [1, n]$,

$$\Sigma_{in} = \sum_{j=1, j\ne i}^n \sigma_{jin}, \quad \sigma_{jin} = \frac{\mathbb{1}\left(X_i \in \mathcal{X}_0\right)}{nb_0^d \widehat{g}_{in}}\varepsilon_j K_0\left(\frac{X_j - X_i}{b_0}\right).$$

Since under (A_4), the σ_{jin}'s, $j \in [1, n]$, are centered independent variables given $X_1, \ldots, X_n$, this yields, for any $k \in \{4, 6\}$,

$$\mathbb{E}_n\left[\Sigma_{in}^k\right] \le C\mathbb{E}\left[\varepsilon^k\right]\left[\frac{\mathbb{1}\left(X_i \in \mathcal{X}_0\right)}{(nb_0^d)^2\widehat{g}_{in}^2}\sum_{j=1}^n K_0^2\left(\frac{X_j - X_i}{b_0}\right)\right]^{k/2} \le \frac{C\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\widetilde{g}_{in}^{k/2}}{(nb_0^d)^{(k/2)}\widehat{g}_{in}^k}.$$

Hence this bound and (A.16) give

$$\mathbb{E}_n\left[\mathbb{1}(X_i \in \mathcal{X}_0)\left(\widehat{m}_{in} - m(X_i)\right)^k\right] \le C\left[\beta_{in}^k + \frac{\mathbb{1}\left(X_i \in \mathcal{X}_0\right)\widetilde{g}_{in}^{k/2}}{(nb_0^d)^{(k/2)}\widehat{g}_{in}^k}\right],$$

which proves (A.15), and then completes the proof of the lemma. □

Proof of Lemma 6.11

Since $K_0(\cdot)$ has a compact support under (A_7), there is a $C > 0$ such that $\|X_i - X_j\| \ge Cb_0$ implies that for any integer number k of $[1, n]$, $K_0((X_k - X_i)/b_0) = 0$ if $K_0((X_j - X_k)/b_0) \ne 0$. Let $D_j \subset [1, n]$ be such that

an integer number k of $[1,n]$ is in D_j if and only if $K_0((X_j - X_k)/b_0) \neq 0$. Abbreviate $\mathbb{P}(\cdot|X_1,\ldots,X_n)$ into $\mathbb{P}_n$ and assume that $\|X_i - X_j\| \geq Cb_0$ so that D_i and D_j have an empty intersection. Note also that taking C large enough ensures that i is not in D_j and j is not in D_i. It then follows, under (A_4) and since D_i and D_j only depend upon $X_1,\ldots,X_n$,

$$
\begin{aligned}
&\mathbb{P}_n\Big((\widehat{m}_{in} - m(X_i), \varepsilon_i) \in A \text{ and } (\widehat{m}_{jn} - m(X_j), \varepsilon_j) \in B\Big) \\
&= \mathbb{P}_n\left(\left(\frac{\sum_{k\in D_i\setminus\{i\}} (m(X_k) - m(X_i) + \varepsilon_k) K_0((X_k - X_i)/b_0)}{\sum_{k\in D_i\setminus\{i\}} K_0((X_k - X_i)/b_0)}, \varepsilon_i\right) \in A\right. \\
&\qquad\qquad \left.\text{and} \left(\frac{\sum_{\ell\in D_j\setminus\{j\}} (m(X_\ell) - m(X_j) + \varepsilon_\ell) K_0((X_\ell - X_j)/b_0)}{\sum_{\ell\in D_j\setminus\{j\}} K_0((X_\ell - X_j)/b_0)}, \varepsilon_j\right) \in B\right) \\
&= \mathbb{P}_n\left(\left(\frac{\sum_{k\in D_i\setminus\{i\}} (m(X_k) - m(X_i) + \varepsilon_k) K_0((X_k - X_i)/b_0)}{\sum_{k\in D_i\setminus\{i\}} K_0((X_k - X_i)/b_0)}, \varepsilon_i\right) \in A\right) \\
&\qquad\qquad \times \mathbb{P}_n\left(\left(\frac{\sum_{\ell\in D_j\setminus\{j\}} (m(X_\ell) - m(X_j) + \varepsilon_\ell) K_0((X_\ell - X_j)/b_0)}{\sum_{\ell\in D_j\setminus\{j\}} K_0((X_\ell - X_j)/b_0)}, \varepsilon_j\right) \in B\right) \\
&= \mathbb{P}_n\left((\widehat{m}_{in} - m(X_i), \varepsilon_i) \in A\right) \times \mathbb{P}_n\left((\widehat{m}_{jn} - m(X_j), \varepsilon_j) \in B\right).
\end{aligned}
$$

This gives the result of Lemma 6.11, since both $(\widehat{m}_{in} - m(X_i), \varepsilon_i)$ and $(\widehat{m}_{jn} - m(X_j), \varepsilon_j)$ are independent given $X_1,\ldots,X_n$. □

Proof of Lemma 6.12

Since $\widehat{m}_{in} - m(X_i)$ depends only upon $(X_1,\ldots,X_n,\varepsilon_k, k\neq i)$, we have

$$
\sum_{i=1}^n \mathrm{Var}_n(\zeta_{in}) \leq \sum_{i=1}^n \mathbb{E}_n\left[\zeta_{in}^2\right] = \sum_{i=1}^n \mathbb{E}_n\left[\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in} - m(X_i))^4\,\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)^2\right]\right],
$$

with, using Lemma 6.7-(6.2),

$$
\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i - \epsilon}{b_1}\right)^2\right] = \int K_1^{(2)}\left(\frac{e-\epsilon}{b_1}\right)^2 f(e)de \leq Cb_1.
$$

Therefore these bounds and Lemma 6.10 give

$$
\begin{aligned}
\sum_{i=1}^n \mathrm{Var}_n(\zeta_{in}) &\leq Cb_1 \sum_{i=1}^n \mathbb{E}_n\left[\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in} - m(X_i))^4\right] \\
&\leq Cnb_1 \sup_{1\leq i\leq n} \mathbb{E}_n\left[\mathbb{1}(X_i\in\mathcal{X}_0)(\widehat{m}_{in} - m(X_i))^4\right] \\
&\leq O_{\mathbb{P}}(nb_1)\left(b_0^4 + \frac{1}{nb_0^d}\right)^2.
\end{aligned}
$$

which yields the desired result for the conditional variance.

We now prepare to compute the order of the conditional covariance. To that aim, observe that Lemma 6.11 gives

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathrm{Cov}_n\left(\zeta_{in},\zeta_{jn}\right)=\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\Big(\mathbb{E}_n\left[\zeta_{in}\zeta_{jn}\right]-\mathbb{E}_n\left[\zeta_{in}\right]\mathbb{E}_n\left[\zeta_{jn}\right]\Big).$$

The order of the term above is derived from the following equalities:

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\mathbb{E}_n\left[\zeta_{in}\right]\mathbb{E}_n\left[\zeta_{jn}\right] = O_{\mathbb{P}}\left(n^2b_0^db_1^6\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2, \tag{A.17}$$

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\mathbb{E}_n\left[\zeta_{in}\zeta_{jn}\right] = O_{\mathbb{P}}\left(n^2b_0^db_1^{7/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2. \tag{A.18}$$

Indeed, since b_1 goes to 0 under (A_{10}), (A.17) and (A.18) yield that

$$\begin{aligned}\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathrm{Cov}_n\left(\zeta_{in},\zeta_{jn}\right) &= O_{\mathbb{P}}\left[\left(n^2b_0^db_1^6\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2+\left(n^2b_0^db_1^{7/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2\right]\\ &= O_{\mathbb{P}}\left(n^2b_0^db_1^{7/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2,\end{aligned}$$

which gives the result for the conditional covariance. Hence, it remains to prove (A.17) and (A.18). For (A.17), note that by (A_4) and Lemma 6.7-(6.2), we have

$$\begin{aligned}\left|\mathbb{E}_n\left[\zeta_{in}\right]\right| &= \left|\mathbb{E}_n\left[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^2\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right]\right|\\ &\leq Cb_1^3\Bigg(\mathbb{E}_n\Big[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^4\Big]\Bigg)^{1/2}.\end{aligned}$$

Hence from this bound and Lemma 6.10 we deduce

$$\begin{aligned}\sup_{1\leq i,j\leq n}\left|\mathbb{E}_n\left[\zeta_{in}\right]\mathbb{E}_n\left[\zeta_{jn}\right]\right| &\leq Cb_1^6\sup_{1\leq i\leq n}\mathbb{E}_n\left[\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^4\right]\\ &\leq O_{\mathbb{P}}\left(b_1^6\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2.\end{aligned}$$

Therefore, since the Markov inequality gives

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)=O_{\mathbb{P}}(n^2b_0^d), \tag{A.19}$$

it then follows that

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\mathbb{E}_n\left[\zeta_{in}\right]\mathbb{E}_n\left[\zeta_{jn}\right]=O_{\mathbb{P}}\left(n^2b_0^db_1^6\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2,$$

which proves (A.17).

For (A.18), set $Z_{in}=\mathbb{1}\left(X_i\in\mathcal{X}_0\right)\left(\widehat{m}_{in}-m(X_i)\right)^2$, and note that for $i\neq j$, we have

$$\mathbb{E}_n\left[\zeta_{in}\zeta_{jn}\right]=\mathbb{E}_n\left[Z_{in}K_1^{(2)}\left(\frac{\varepsilon_j-\epsilon}{b_1}\right)\mathbb{E}_{in}\left[Z_{jn}K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right], \tag{A.20}$$

where

$$\begin{aligned}&\mathbb{E}_{in}\left[Z_{jn}K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\\ &=\ \beta_{jn}^2\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]+2\beta_{jn}\mathbb{E}_{in}\left[\Sigma_{jn}K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]+\mathbb{E}_{in}\left[\Sigma_{jn}^2K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right].\end{aligned} \tag{A.21}$$

The first term of Equality (A.21) is treated by using Lemma 6.7-(6.2). This gives

$$\left|\beta_{jn}^2\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right|\leq Cb_1^3\beta_{jn}^2. \tag{A.22}$$

Since under (A_4), the ε_j's are independent centered variables, and are independent of the X_j's, the second term in (A.21) gives

$$\begin{aligned}\mathbb{E}_{in}\left[\Sigma_{jn}K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]&=\frac{\mathbb{1}\left(X_j\in\mathcal{X}_0\right)}{nb_0^d\widehat{g}_{jn}}\sum_{k=1,k\neq j}^{n}K_0\left(\frac{X_k-X_j}{b_0}\right)\mathbb{E}_{in}\left[\varepsilon_kK_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\\ &=\frac{\mathbb{1}\left(X_j\in\mathcal{X}_0\right)}{nb_0^d\widehat{g}_{jn}}K_0\left(\frac{X_i-X_j}{b_0}\right)\mathbb{E}_{in}\left[\varepsilon_iK_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right].\end{aligned}$$

Therefore, by (A_7) which ensures that K_0 is bounded, the equality above and Lemma 6.7-(6.2) yield that

$$\left|\beta_{jn}\mathbb{E}_{in}\left[\Sigma_{jn}K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right|\leq Cb_1^3\left|\beta_{jn}\frac{\mathbb{1}\left(X_j\in\mathcal{X}_0\right)}{nb_0^d\widehat{g}_{jn}}\right|. \tag{A.23}$$

For the last term in (A.21), we have

$$\begin{aligned}&\mathbb{E}_{in}\left[\Sigma_{jn}^2(x)K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\\ &=\ \frac{1}{(nb_0^d\widehat{g}_{jn})^2}\sum_{\substack{k=1\\k\neq j}}^{n}\sum_{\substack{\ell=1\\\ell\neq j}}^{n}K_0\left(\frac{X_k-X_j}{b_0}\right)K_0\left(\frac{X_\ell-X_j}{b_0}\right)\mathbb{E}_{in}\left[\varepsilon_k\varepsilon_\ell K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\\ &=\ \frac{1}{(nb_0^d\widehat{g}_{jn})^2}\sum_{k=1,k\neq j}^{n}K_0^2\left(\frac{X_k-X_j}{b_0}\right)\mathbb{E}_{in}\left[\varepsilon_k^2K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right],\end{aligned}$$

with, using Lemma 6.7-(6.2),

$$
\begin{aligned}
&\left|\mathbb{E}_{in}\left[\varepsilon_k^2 K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right| \\
&\quad \leq \max\left\{\sup_{e\in\mathbb{R}}\left|\mathbb{E}_{in}\left[\varepsilon^2 K_1^{(2)}\left(\frac{\varepsilon-e}{b_1}\right)\right]\right|,\ \mathbb{E}[\varepsilon^2]\sup_{e\in\mathbb{R}}\left|\mathbb{E}_{in}\left[K_1^{(2)}\left(\frac{\varepsilon-e}{b_1}\right)\right]\right|\right\} \\
&\quad \leq Cb_1^3.
\end{aligned}
$$

Therefore

$$
\left|\mathbb{E}_{in}\left[\Sigma_{jn}^2 K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right| \leq \frac{Cb_1^3}{(nb_0^d\widehat{g}_{jn})^2}\sum_{k=1,k\neq j}^{n} K_0^2\left(\frac{X_k-X_j}{b_0}\right).
$$

Substituting this bound, (A.23) and (A.22) in (A.21), we obtain

$$
\left|\mathbb{E}_{in}\left[Z_{jn}K_1^{(2)}\left(\frac{\varepsilon_i-\epsilon}{b_1}\right)\right]\right| \leq Cb_1^3 M_n,
$$

where

$$
M_n = \sup_{1\leq j\leq n}\left[\beta_{jn}^2 + \left|\beta_{jn}\frac{\mathbb{1}\left(X_j\in\mathcal{X}_0\right)}{nb_0^d\widehat{g}_{jn}}\right| + \frac{1}{(nb_0^d\widehat{g}_{jn})^2}\sum_{k=1,k\neq j}^{n} K_0^2\left(\frac{X_k-X_j}{b_0}\right)\right].
$$

Hence from (A.20), the Cauchy-Schwarz inequality, Lemma 6.10 and Lemma 6.7-(6.2), we deduce

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\left|\mathbb{E}_n\left[\zeta_{in}\zeta_{jn}\right]\right| \\
&\quad \leq CM_n b_1^3\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\mathbb{E}_n\left|Z_{in}K_1^{(2)}\left(\frac{\varepsilon_j-\epsilon}{b_1}\right)\right| \\
&\quad \leq CM_n b_1^3\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\mathbb{E}_n^{1/2}\left[Z_{in}^2\right]\mathbb{E}_n^{1/2}\left[K_1^{(2)}\left(\frac{\varepsilon_j-\epsilon}{b_1}\right)^2\right] \\
&\quad \leq M_n b_1^3 O_{\mathbb{P}}\left(b_0^4+\frac{1}{nb_0^d}\right)(b_1)^{1/2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\Big(\mathbb{1}\left(\|X_i-X_j\|\leq Cb_0\right)\Big).
\end{aligned}
$$

Moreover, (A.8) and Lemma 6.1 give, under (A_1), (A_7) and (A_9),

$$
M_n = O_{\mathbb{P}}\left(b_0^4+\frac{b_0^2}{nb_0^d}+\frac{1}{nb_0^d}\right) = O_{\mathbb{P}}\left(b_0^4+\frac{1}{nb_0^d}\right).
$$

Finally, substituting this order in the bound above, and using (A.19), we arrive at

$$
\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\mathbb{1}\Big(\|X_i-X_j\|<Cb_0\Big)\mathbb{E}_n\left[\zeta_{in}\zeta_{jn}\right] = O_{\mathbb{P}}\left(n^2 b_0^d b_1^{7/2}\right)\left(b_0^4+\frac{1}{nb_0^d}\right)^2.
$$

This proves (A.18), and then completes the proof of the theorem. □

References

[1] Ahmad, I. & Li, Q. (1997). Testing symmetry of an unknown density function by kernel method. *Nonparam. Statistics.* **7**, 279–293.

[2] Altman, N. and Léger, C. (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plann. Inf.* **46**, n°2, 195–214.

[3] Akritas, M. G. & Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scandinavian Journal of Statistics.* **28**, 549–567.

[4] Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika.* **68**, n°1, 326-328.

[5] Bai, J. (1994). Weak convergence of the sequential empirical processes in ARMA models. *Ann. Statist.* **22**, 2051–2061.

[6] Bashtannyk, D., M. and Hyndman, R., J. (2001). Bandwidth selection for conditional density estimation. *Comput. Statist. Data Anal.* **36**, n°3, 279–298.

[7] Bertrand-Retali, M. (1978). Convergence uniforme d'un estimateur de la densité par la méthode du noyau. *Rev. Roumaine Math. Pures et Appliquées.* **23**, n°3, 361-385.

[8] G. Biau, B. Cadre, B. Pelletier, D. Mason. (2009). Asymptotic normality in density support estimation. *Electron. J. Probab.*, 2617-2635.

[9] Biau, G., Cadre, B. and Pelletier, B. (2008). Exact rates in density support estimation. *J. Multivariate Anal.* **99**, 2185–2207.

[10] Bickel, P.J., Klaasssen, C.A.J., Ritov, Y. and Wellner, J.A. (1998). *Efficient and adaptive estimation for semi-parametric models.* Springer, New York.

[11] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley.

[12] Boldin, M. V. (1982). Estimation of the distribution of noise in an autoregressive scheme. *Theory Probab. Appl.* **27**, 866–871.

[13] Boldin, M. V. (1998). On residual empirical distribution functions in ARCH models with applications to testing and estimation. *Mitt. Math. Sem. Giessen* **235**, 49–66.

[14] Bosq, D. and Blanke, D. (2007). *Inference and prediction in large dimensions.* Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester Economica.

[15] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction.* Lecture Notes in Staistics, 110, 2nd ed. Springer-Verlag, New York.

[16] Bosq, D., Lecoutre, J.P. (1987). *Théorie de l'estimation fonctionnelle.* Economica.

[17] Borovkov, A. (1987). *Statistique mathématique.* MIR, Moscou.

[18] Bowman, A., Hall, P. and Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika.* **85**, n°4, 799-808.

[19] Cadre, B. (2006). Kernel estimation of density levels sets. *J. Multiv. Anal.* **97**, n°4, 999–1023.

[20] Breimen, L., Meisel, W. and Purcell, E. (1977). Variable kernel estimate of multivariate densities. *Technometrics.* **19**, 135–144.

[21] Cheng, F. (2005). Asymptotic distributions of error density and distribution function estimators in nonparametric regression. *Journal of Statistical Planning and Inference* **128**, 327–349.

[22] Cheng, F., Sun, S. (2008). A goodness-of-fit test of the errors in nonlinear autoregressive time series models. *Statistics and Probability Letters* **78**, 50–59.

[23] Chow, Y., S. & Teicher, H. (2003). *Probability Theory: Independence, Interchangeability, Martingales.* Springer, 3rd ed.

[24] Conover, W., J. (1999). *Pratical Nonparametric Statistics.*, Third Edition, John Wliey & Sons, New York.

[25] Dallal, G., E. and Wilkinson, L. (1986). An analytic approximation to the distribution of Lilliefors' test statistic for normality. *The American Statistician.* **40**, 294–296.

[26] Collomb, G. (1979). Conditions nécessaires et suffisantes de convergence uniforme d'un estimator de régression, estimation des dérivées de la régression. *C.R. Acad. Sci. Paris Serie A.* **288**, 161–163.

[27] D'Agostino, R. and Stephens, M. (1986). *Goodness of fit techniques.*, M. Dekker, Inc., New York.

[28] De Gooijer, J., G. and Zerom, D. (2003). On conditional density estimation. *Statist. Neerlandica.* **57**, n°2, 159–176.

[29] Denby, L., Martin, D. (1979). Robust estimation of the first order autoregressive parameter. *J. Amer. Statist. Assoc.* **74**, 140–146.

[30] Dette, H. and von Lieres und Wilkau, C. (2001). Testing additivity by kernel based methods–What is a reasonable test? *Bernoulli* **7**, 669–697.

[31] Dette, H., Kusi-Appiah, S., Neumeyer, N., (2002). Testing symmetry in nonparametric regression models. *Nonparam. Statistics.* **14(5)**, 477–494.

[32] Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Annals of Statistics.* **1)**, 279–290.

[33] Efromovich, S. (2005). Estimation of the density of the regression errors. *Annals of Statistics.* **33**, n°5, 2194–2227.

[34] Efromovich, S. (2007). Adaptive estimation of error density in nonparametric regression with small sample size. *Journal of Statistical Planning and Inference.* **137**, 363–378.

[35] Einmahl, U. & Mason, D. M. (2005). Uniform in bandwidth consistency of Kernel -type functions estimators. *Annals of Statistics.* **33**, 1380–1403.

[36] Escanciano, J., C. and Jacho-Chavez, D. (2010). $\sqrt{n}$-uniformly consistent density estimation in nonparametric regression models (submitted).

Einmahl, U. & Mason, D. M. (2005). Uniform in bandwidth consistency of Kernel -type functions estimators. *Annals of Statistics.* **33**, 1380–1403.

[37] Fan, J. and Yim, T., H. (2004). A crossvalidation method for estimating conditional densities. *Biometrika.* **91**, n° 4, 819–834.

[38] Feller, W. (1948). On the Kolmogorov-Smirnov Limit Theorems for Empirical Distributions. *Annals of Mathematical Statistics.* **19**, 177–189.

[39] Ferignac, P. (1962). Test de Kolmogorov-Smirnov sur la validité d'une fonction de distribution. *Revue de Statistique Appliquée.* **tome 10**, n° 4, 13–32.

[40] Fu, K., Yang, X. (2008). Asymptotics of Kernel error density estimators in nonlinear autoregressive models. *J. Math. Chem.* **44**, 831–838.

[41] Francq, C. and Tran, L., T. (2002). Nonparametric estimation of density, regression and dependence coefficients. *Journal of Nonparametric Statistics.* **14**, n°6, 729–747.

[42] Gibbons, J., D. and Chakraborti, S. (2003). *Nonparametric Statistical Inference.*, Fourth Edition, Revised and Expanded, Marcel Dekker, Inc.

[43] Giné, E. and Guillou A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. Henry Poincaré Prob. Statistics.* **38**, 907–921.

[44] Gozalo, P., L. and Linton, O., B. (2001). A nonparametric test of additivity in generalized nonparametric regression with estimated parameters. *J. Econometrics.* **104**, 1–48.

[45] Györfi, L. and Kohler, M. (2007). Nonparametric estimation of conditional distributions. *IEEE Trans. Inform. Theory.* **53**, n°5, 1872–1879.

[46] Hall, P., Racine, J. and Li, Q. (2004). Cross-validation and the estimation of conditional probability densities. *J. Amer. Statist. Assoc.* **99**, n° 468, 1015–1026.

[47] Hall, P., Carroll, R.J (1989). Variance function estimation in regression: The effect of estimating mean. *J. Roy. Statist. Soc. Ser. B,* **51**, 3–14.

[48] Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and semiparametric models.* Springer Series in Statistics. Springer-Verlag, New-York.

[49] Hjort, N., L. and Jones, M., C. (1996). Locally parametric nonparametric density estimation. *Ann. Statist.* **24**, n°4, 1619–1647.

[50] Hössjer, O. and Ruppert, D. (1995). Asymptotics for the transformation kernel density estimator. *Ann. Statist.* **23**, n°4, 1198–1222.

[51] Huber, P., H. (1981). Robust estimation of a location parameter. *Annals Math. Statist.* **35**, 73–101.

[52] Huber, P., H. (1981). *Robust Statistics.* Wiley, New York.

[53] Hyndman, R., J. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. *J. Nonpar. Statist.* **14**, n°3, 259–278.

[54] Hyndman, R., J., Bashtannyk, D.,M. and Grunwald, G., K. (1996). Estimating and visualizing conditional densities. *J. Comput. Graph. Statist.* **5**, n°4, 315–336.

[55] Jones, M.C., Marron, J.S. and Sheather., S.J. (1996). A brief survey of bandwidth selection for density estimation. *J. Amer. Statist. Assoc.* **91**, 401–407.

[56] Klimko, L.A., Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6**, 629–642.

[57] Kolmogorov, A. (1933) Grundbegriffe der Wahrscheinlichkeitsrechnung. Berlin: Julius Springer. Translation: Kolmogorov, Andrey (1956) in Foundations of the Theory of Probability, 2nd ed., New York: Chelsea. pp. 50–51.

[58] Koul, H. L. (1992). *Weighted empiricals and linear models.* IMS, Hayward, CA.

[59] Koul, H. L., Sakhanenko, L. (2005). Goodness-of-fit testing in regression: A finite sample comparison of boostrap methodology and Khmaladze transformation. *Statistics and Probability Letters.* **74**, 290–302.

[60] Lehmann, E. L. and Romano, J., P. (1998). *Testing Statistical Hypotheses.*, Third Edition, Springer.

[61] Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Computational Statistics and Data Analysis.* **14**, n° 4, 457–471.

[62] Lepski, O.V. and Spokoiny, V.G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Annals of Statistics* **25**, 2512–2546.

[63] Loftsgaarden, D.O. and Quesenberry, C.P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.* **36**, 1049–1051.

[64] Loynes, R.M., (1980). The empirical distribution function of residuals from generalised regression. *Annals of Statistics* **8**, 285–298.

[65] Mammen, M. (1996). Empirical process of residuals for high-dimensional linear models. *Ann. Statist.* **24**, 307–335.

[66] Massey, F., J., Jr (1951). The Komogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association.* **46**, 253 (March 1951), 68–78.

[67] Miller, L., H. (1956). Table of percentage points of Kolmogorov Statistics. *Journal of the American Statistical Association.* **51**, 273 (March 1956), 111–121.

[68] Müller, U. U., Schick, A. et Wefelmeyer, W. (2004). Estimating linear functionals of the error distribution in nonparametric regression. *J. Statist. Plann. Inference.* **119**, 75–93.

[69] Müller, H.G. and Stadmüller, U. (1987). Variable bandwidth kernel estimators of regression curves. *Annals of Statistics* **15**, 182–201.

[70] Nadaraya, E., A. (1964). On a regression estimate. *Teor. Verojatnost. i Primenen.* **9**, 157–159.

[71] Nadaraya, E. A. (1989). *Nonparametric estimation of probability densities and regression curves.* Kluwer Academic Publishers.

[72] Neumeyer, N., Dette, H., Nagel, E-R. (2005). A note on testing symmetry of the error distribution in linear regression models. *Nonparametric Statistics* **17**, n°6, 697–715.

[73] Neumeyer, N. and Dette, H. (2007). Testing for symmetric error distribution in nonparametric regression models. *Statistica Sinica* **17**, 775–795.

[74] Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *J. Multiv. Analysis.* **101**, 1067–1078.

[75] Parzen, E. (1962). On estimation of a probability distribution function and mode. *Ann. Math. Statist.* **33**, 1065–1076.

[76] Pinsker, M.S. (1980). Optimal filtering of a square integrable signal in Gaussian white noise. *Problems Inform. Transmission* **16**, 52–68.

[77] Plancade, S. (2008). Nonparmetric estimation of the density of the regression noise. *C.R. Acad. Sci. Paris* **Ser. I 346**, 461–466.

[78] Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

[79] Portnoy, S. (1986). Asymptotic behavior of the empirical distribution of M-estimated residuals from a regression model with many parameters. *Ann. Statist.,* **14**, 1152–1170.

[80] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of density function. *Ann. Math. Statist.,* **27**, 832–837.

[81] Roussas, G. (1967). Nonparametric Estimation in Markov processes. *Technical Repport* **110**, *Univ. of Wisconsin, Madison.*

[82] Roussas, G. (1991). Estimation of transition distribution function and its quantiles in Markov processes: strong consistency and asymptotic normality. In: *Nonparametric Functional Estimation and Related Topics*, pp. 463-474. *Kluwer, Dordrecht.*

[83] Sarda, P., Vieu, P. and Youndjé, E. (1996). Choix de paramètres de lissage en estimation de densité conditionnelle. *Publi. de l'Inst. de Statist. de Paris,* **XXXVIII**, 57–80.

[84] Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions. *J. Statist. Plann. Inference,* **35**, n°1, 65–75.

[85] Scott, W., S. (1992). *Multivariate density estimation.* Wiley.

[86] Serfling, R.J. (2002). *Approximation Theorems of Mathematical Statistics.* Paperback Edition, Wiley.

[87] Shorack, G. R. (1984). Empirical and rank processes of observations and residuals. *Canad. J. Statist.,* **12**, 319–332.

[88] Shorack, G.,R. and Wellner, J., A. (1986).*Empirical Processes With Applications to Statistics*, John Wiley & Sons Inc.

[89] Smirnov, N. (1948). Table for estimating goodness of fit of empirical distributions. *Annals of Mathematical Statistics,* **19**, 279–281.

[90] Staniswalis, J. G. (1989). Local bandwidth selection for kernel estimates. *Journal of American Statistical Association,* **84**, 284–288.

[91] Stute, W. (1982). The law of the iterated logarithm for kernel density estimators. *Annals of Probability,* **10**, 414–422.

[92] Stone, C. J. (1977). Consistent nonparametric regression. *Annals of Statistics,* **5**, 595–645.

[93] Tassi, P. (1985). *Méthodes Statistiques.*, Ed. Economica.

[94] Terrell, G.R. and Scott, D.W. (1992). Variable kernel density estimation. *Annals of Statistics,* **20**, n°3, 1236–1265.

[95] Thode Jr, H., C. (2002). *Testing for normality.*, Marcel Dekker, New York.

[96] Van Der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge Univ. Press.

[97] Van Der Vaart, A. W., Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer, New York.

[98] Van Keilegom, I. and Veraverbeke, N. (2002). Density and hazard estimation in censored regression models. *Bernoulli,* **8(5)**, 607–625.

Watson, G., S. (1964). Smooth regression analysis. *Sankhyā*, **Ser. A 26**, 359–372.

[99] Wand, M., P., Jones, M., C. (1995). *Kernel Smoothing.* Chapman & Hall/CRC.

[100] Wang, L., Brown, L. D., Cai, T. T., Levine, M. (2008). Effect of mean on variance function estimation in nonparametric regression. *Ann. Statist.,* **36**, 646–664.

[101] Watson, G., S. (1964). Smooth regression analysis. *Sankhyā*, **Ser. A 26**, 359–372.

[102] Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability and its Applications,* **5**, 302–305.

[103] Youndjé, E. (1996). Propriétés de convergence de l'estimateur à noyau de la densité conditionnelle. *Revue Roumaine de Mathématiques Pures et Appliquées,* **41**, 535–566.

[104] Youndjé, E., Sarda, P. and Vieu, P. (1994). Validation croisée pour l'estimation non-paramétrique de la densité conditionnelle. *Publ. Inst. Statist. Univ. Paris,* **38**, n°1, 57–80.